

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 786 519 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.07.1997 Bulletin 1997/31

(51) Int Cl.⁶: **C12N 15/00**

(21) Application number: **97100117.7**

(22) Date of filing: **07.01.1997**

(84) Designated Contracting States:
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**
Designated Extension States:
AL LT LV RO SI

(30) Priority: **05.01.1996 US 9861**

(71) Applicant: **HUMAN GENOME SCIENCES, INC.**
Rockville, MD 20850-3338 (US)

(72) Inventors:
• **Kunsch, Charles A.**
Gaithersburg, Maryland 20882 (US)

- **Choi, Gil H.**
Rockville, Maryland 20850 (US)
- **Barash, Steven C.**
Rockville, Maryland 20850 (US)
- **Dillon, Patrick J.**
Gaithersburg, Maryland 20879 (US)
- **Fannon, Michael R.**
Silver Spring, Maryland 20906 (US)
- **Rosen, Craig A.**
Laytonsville, Maryland 20882 (US)

(74) Representative: **VOSSIUS & PARTNER**
Postfach 86 07 67
81634 München (DE)

(54) **Staphylococcus aureus polynucleotides and sequences**

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu-

cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

EP 0 786 519 A2

Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., *The Staphylococcus* as a Molecular Genetic System, Chapter 1, pgs. 1-37 in *MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI*, R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

Human Health and *S. Aureus*

Staphylococcus aureus is a ubiquitous pathogen. (See, for instance, Mims *et al.*, *MEDICAL MICROBIOLOGY*, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyrogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic streptococci. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

Eyelid infections

S. aureus is the cause of styes and of sticky eye" in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

Joint infections

S. aureus infects bone joints causing diseases such osteomyelitis.

Osteomyelitis

S. aureus is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphyseal growth plates in the end of long, growing bones.

Skin infections

S. aureus is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*.

Surgical Wound Infections

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. *S. aureus* is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer *S. aureus* cells than are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

Scalded Skin Syndrome

S. aureus is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This disease occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains produce exfoliation (also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the disease. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

Toxic Shock Syndrome

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxemia and septicaemia, and can be fatal.

Nocosomal Infections

In the 1984 National Nosocomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

Resistance to drugs of S. aureus strains

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to penicilloic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance; however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, strains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI, R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J. , *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs," fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs," and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vaccinating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs), expression modulating fragment (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC®).

The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* **85**: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; i.e., a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (e.g., a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* **215**:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* **17**:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences):

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

5 BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which
10 modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification
15 means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-5,191, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These
20 include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well
25 known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of *Staphylococcus aureus* genomic DNA. Thus, given the availability of SEQ ID NOS:1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were
35 identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the Reference* for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name* of the matching sequence; column eight provides the BLAST identity* score from the comparison of the ORF and the homologous gene; and column nine
55 indicates the length in nucleotides of the highest scoring segment pair* identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (e.g., at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5,191, with a sequence from another isolate of the same species.

Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*

phylococcus aureus, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6: 3073 (1979); Cooney *et al.*, *Science* 241: 456 (1988); and Dervan *et al.*, *Science* 251: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus*, *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Others

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes *Staphylococcus aureus* polypeptides which are useful as immunodiagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outer membrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., *Mol. Microbiol.* 13, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* 174, 7345-7351; 1992)). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. Lipoproteins in bacteria. *J. Bioenerg. Biomembr.* 22, 451-471; 1990).

It well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. *ASM News* 62, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

aminated. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen".

5 An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outermembrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides
10 listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF
15 *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic
20 amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding
25 ORF listed in Tables 1,2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of
30 amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for
40 instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Leamer, R. A. (1983) "Antibodies that react with
45 predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance,
50 Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *S. aureus* specific antibodies include: a polypeptide comprising peptides shown in Table
55 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated *S. aureus* proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a per-alkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among these are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis et al., PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C in 6X SSPE and 50% formamide, and washing at 50-65°C in 0.5X SSPE) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Staphylococcus aureus*.

ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESIS, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389: 93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872: 83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B- cell hybridoma technique (Kozbor *et al.*, *Immunology Today*

4: 72 (1983), pgs. 77-96 of Cole *et al.*, in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W. J. *Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs, antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry*; PCT publication WO95/32291, and *Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the Dfs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed Dfs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Staphylococcus aureus* genome; and
- (b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6:3073 (1979); Cooney *et al.*, Science 241:456 (1988); and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic *Staphylococcus aureus* polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLES**LIBRARIES AND SEQUENCING****1. Shotgun Sequencing Probability Analysis**

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P_0 , that any given base in a sequence of size L , in nucleotides, is not sequenced after a certain amount, n , in nucleotides, of random sequence has been determined can be calculated by the equation $P_0 = e^{-m}$, where m is L/n , the fold coverage." For instance, for a genome of 2.8 Mb, $m=1$ when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P_0 = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G , is determined by the equation $G = Le^{-m}$, and the average gap size, g , follows the equation, $g = L/n$. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Staphylococcus aureus DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer. The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37° C in a reaction mixture (50 ul) containing the v+i linears, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 µl aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl₂ (1 M), and 1 ml MgSO₄/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 µl aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 µl) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucrose gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation. One µl of fragments was used with 1 µl of DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500 µl of recommended SM buffer and chloroform treatment). Yield was about 2.5x10⁹ pfu/µl.

An amplified library was prepared from the primary packaging mixture according to the manufacturer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x10⁹ pfu/ml.

Mini-liquid lysates (0.1 µl) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

INFORMATICS

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

3. Identifying Genes

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of *Staphylococcal* disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic reagent.

3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

4. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and Sall restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternatively and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* ExpressTM Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	1419	757	[emb]X17301 SAILD	S. aureus DNA for hld gene and for part of agr gene	100	663	663
1	2	3273	2452	[emb]X52543 SAAG	S. aureus agrA, agrB and hld genes	99	809	822
1	5	6418	5651	[dbj]D14711 STAM	Staphylococcus aureus HSP10 and HSP60 genes	98	223	768
5	1	807	439	[emb]X72700 SAPV	S. aureus genes for S and F components of Panton-Valentine leucocidins	81	216	369
5	4	5031	3571	[emb]X72700 SAPV	S. aureus genes for S and F components of Panton-Valentine leucocidins	95	424	1461
10	1	86	904	[gb]L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grIA and grIB) genes, complete cds	98	715	819
16	5	5302	6246	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	94	251	945
16	6	6249	7091	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	99	843	843
16	7	7084	7584	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	99	342	501
20	1	995	549	[gb]L19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	443	447
20	2	1011	841	[gb]L19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	91	137	171
20	3	2010	1798	[gb]L19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	110	213
20	4	5300	3825	[gb]M16714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	948	1476
20	5	4788	4282	[gb]M16714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	309	507
26	1	2	145	[gb]U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (lles) gene, partial cds	100	126	144
26	2	84	557	[gb]U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (lles) gene, partial cds	99	430	474
26	3	763	3531	[emb]X74219 SAIL	S. aureus gene for isoleucyl-tRNA synthetase	99	2769	2769
29	3	1261	4392	[gb]U66665	Staphylococcus aureus DNA fragment with class II promoter activity	100	117	3132
31	14	14977	13463	[emb]X73889 SAP1	S. aureus genes P1 and P2	99	1351	1515
31	15	14241	13855	[emb]X73889 SAP1	S. aureus genes P1 and P2	98	258	387
38	17	14284	13112	[gb]M12715	S. aureus gah gene encoding lipase (glycerol ester hydrolase)	100	372	1173
38	19	13434	15518	[gb]M12715	S. aureus gah gene encoding lipase (glycerol ester hydrolase)	100	2085	2085

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
46	2	519	1727	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8D, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1209	1209
46	3	1720	2295	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	576	576
46	4	2259	3182	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	924	924
46	5	3173	4498	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1283	1326
46	6	4536	5720	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1185	1185
46	7	6455	6120	[gb U73374]	Staphylococcus aureus type 8 capsulo genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	278	336
48	1	2	955	[gb L25893]	Staphylococcus aureus recA gene, complete cds	99	954	954
50	3	4465	2924	[emb X85029]SAAH	S. aureus AhpC gene	100	88	1542
50	4	4108	3515	[emb X85029]SAAH	S. aureus AhpC gene	98	540	594
54	3	5074	3192	[emb X62992]SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	1668	1683
54	4	4865	4122	[emb X62992]SAFN	S. aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	[emb X62992]SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	11386	8300	[gb J04151]	S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds	100	3087	3087
58	3	1743	2819	[emb X87104]SADN	S. aureus mdr, pbp4 and tagD genes (SC511-55 isolate)	89	68	1077
58	4	2858	3280	[emb X91786]SAPB	S. aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	[emb X91786]SAPB	S. aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	[gb U29478]	Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds	100	300	300
58	7	5086	6840	[emb X91786]SAPB	S. aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	[gb 121854]	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	[emb X52543]SAG	S. aureus agrA, agrB and hld genes	99	673	1005

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
82	1	357	3917	emb X64172 SARP	S.aureus rplL, orf202, rps3(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	2396	3561
82	2	4027	7677	emb X89233 SARP	S.aureus DNA for rpoC gene	99	3171	3651
82	3	7745	8068	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	320	324
82	4	8103	8579	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	477	477
82	5	8618	8821	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	154	204
84	1	18	191	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	164	174
84	2	189	893	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705
84	3	887	1660	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774
84	4	1584	3503	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920
84	5	3394	4521	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128
84	6	4519	5643	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125
96	2	1245	3896	emb 218852 SACF	S.aureus gene for clumping factor	83	660	2652
97	2	625	882	gb U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	97	68	258
111	1	3	452	gb L41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	450	450
111	2	526	1041	gb L41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	516	516
117	2	1278	1958	gb M83994	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	61	681

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
118	4	3787	4254	[dbj U0690 STAN cds]	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP10; ORF35, complete	99	467	468
130	4	2597	3640	[emb X13290 SATN]	Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing transposon Tn4003	78	936	1044
130	5	3813	4265	[emb Z16422 SADI]	S.aureus dfrB gene for dihydrofolate reductase	98	416	433
130	6	4309	5172	[emb Z16422 SADI]	S.aureus dfrB gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	[emb X71437 SAGY]	S.aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	[dbj O10489 STAG]	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12886	10940	[dbj U10489 STAG]	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12592	11765	[gb S77055]	recF cluster: dnaA-rep1osome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]	99	822	828
143	3	4171	2867	[gb U36379]	Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds	99	1305	1305
143	4	3100	4281	[gb L42943]	Staphylococcus aureus (clone KIN50) phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	1170	1182
143	5	4254	4718	[gb U51133]	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	449	465
143	9	6977	7261	[gb U51132]	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (aene), and o-succinylbenzoic acid synthetase (aene) genes, complete cds	100	75	285
143	10	9464	8361	[gb U51132]	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (aene), and o-succinylbenzoic acid synthetase (aene) genes, complete cds	100	1108	1104
143	11	11232	9748	[gb U51132]	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (aene), and o-succinylbenzoic acid synthetase (aene) genes, complete cds	100	1485	1485
143	12	10739	10320	[gb U51132]	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (aene), and o-succinylbenzoic acid synthetase (aene) genes, complete cds	100	332	420
152	5	2454	3437	[emb X58434 SAPD]	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	305	984
152	6	3513	4820	[emb X58434 SAPD]	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	1308	1308
152	7	4818	6230	[emb X58434 SAPD]	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	1413	1413
153	1	387	1526	[gb S77055]	recF cluster: dnaA-rep1osome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]	99	1140	1140
153	2	1877	2152	[gb S77055]	recF cluster: dnaA-rep1osome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]	100	276	276

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
153	3	2143	2289	[gb S77055]	[ecf cluster: dnaA replisome assembly protein...gyrB-DNA gyrase beta subunit (Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt)]	99	113	147
154	10	10792	9314	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	91	154	1479
154	11	9935	9615	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	229	321
154	12	9943	10167	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	94	123	225
154	13	10089	11501	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	1326	1413
159	2	2195	1212	[dbj D28879]	[STAP Staphylococcus aureus gene for penicillin-binding protein 1, complete cds]	100	71	984
161	3	2596	2270	[gb H83994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	92	203	327
162	1	1406	705	[gb U21221]	[Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds]	100	702	702
163	4	1263	1772	[gb U19770]	[Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds]	96	127	510
164	7	4774	9117	[dbj D66727]	[D867 Staphylococcus aureus DNA for DNA polymerase III, complete cds]	99	3470	4344
168	7	7448	6447	[gb U21636]	[Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	100	1002	1002
168	8	9538	7961	[gb U21636]	[Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	99	1158	1578
173	6	9240	7801	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	1440	1440
173	7	11252	9522	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	99	1731	1731
173	8	8285	8704	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	420	420
173	9	10168	9839	[gb J03479]	[S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	330	330
173	10	11815	10829	[emb X14827]	[SALA Staphylococcus aureus lacC and lacD genes]	100	987	987
173	11	12721	11774	[emb X14827]	[SALA Staphylococcus aureus lacC and lacD genes]	100	948	948
173	12	12838	12305	[gb M64724]	[S. aureus tagatase 6-phosphate isomerase gene, complete cds]	100	534	534
173	13	13243	12773	[gb U32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds]	100	471	471
173	14	14633	13866	[gb U32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds]	100	768	768

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
178	1	2	655	gb U52961	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	115	654
178	2	2201	1482	gb U52961	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	720	720
178	3	2361	1909	gb U52961	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	453	453
178	4	1551	1853	gb U52961	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	303	303
178	5	3541	2777	gb U52961	Staphylococcus aureus lytS and lytR genes, complete cds	99	765	765
178	6	3294	3025	gb U52961	Staphylococcus aureus lytS and lytR genes, complete cds	99	270	270
181	1	1114	590	gb H61177	S. aureus sigma factor (plac) gene, complete cds	99	499	525
182	1	3	341	emb X61307 SASP	Staphylococcus aureus spa gene for protein A	98	277	339
182	2	690	2312	gb H01786	S. aureus spa gene coding for protein A, complete cds	97	1332	1623
182	3	5861	4251	emb X61307 SASP	Staphylococcus aureus spa gene for protein A	99	119	1611
185	1	3	824	gb U31979	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	90	132	822
191	3	841	2760	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	99	1920	1920
191	4	2967	3143	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	99	177	177
191	5	5768	4566	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	99	250	1203
196	1	1741	872	gb U36472	Staphylococcus aureus lysyl-tRNA synthetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene	99	870	870
198	3	1688	2011	emb X93205 SAPT	S. aureus ptai and ptai genes	99	324	324
198	4	2005	2310	emb X93205 SAPT	S. aureus ptai and ptai genes	97	304	306
202	1	163	1305	emb X97985 SA12	S. aureus orf1, 2, 3 & 4	99	1143	1143
202	2	1303	2175	emb X73889 SAP1	S. aureus genes P1 and P2	94	444	873
210	1	3114	1558	dbj D17386 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1552	1557
210	2	2939	2232	gb U41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	684	708
214	11	7429	7770	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	96	157	342

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
216	3	398	1318	emb X72700 SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidin	88	265	921
219	2	1810	1073	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	100	60	738
219	3	2979	2035	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	945	945
219	4	4359	3196	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	1164	1164
219	5	7044	1116	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	1869	1869
219	6	6557	5883	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	675	675
219	7	6801	6334	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	468	468
221	8	10816	10034	gb U19298	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic) gene, complete cds	91	67	783
223	1	2855	1506	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	102	1350
234	1	2	1357	emb X97985 SA12	S.aureus orfs 1,2,3 & 4	100	176	1356
234	2	1694	2485	emb X97985 SA12	S.aureus orfs 1,2,3 & 4	100	792	792
234	3	2648	3148	emb X97985 SA12	S.aureus orfs 1,2,3 & 4	99	501	501
234	4	3120	4604	emb X97985 SA12	S.aureus orfs 1,2,3 & 4	99	1305	1485
236	6	3826	5322	gb U48826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	96	648	1497
248	1	2	403	emb X62288 SAPV	S.aureus DNA for penicillin-binding protein 2	100	103	402
248	2	388	852	gb U25426	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	465	465
253	2	1539	1093	gb U46541	Staphylococcus aureus sara gene, complete cds	96	447	447
254	2	150	1835	gb U57060	Staphylococcus aureus scdA gene, complete cds	94	142	1686
254	3	1973	2728	gb U57060	Staphylococcus aureus scdA gene, complete cds	99	756	756
260	1	2	1900	gb X90693	Staphylococcus aureus glycerol ester hydrolase (lip) gene, complete cds	99	1213	1899
265	1	1	942	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	ORF nt length
265	2	688	476	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	213
265	3	2418	1765	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	98	654
266	1	2	1018	dbj D14711 STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	743
282	1	1	525	gb F72488	hemB-porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	110
282	2	516	1502	gb S72488	hemB-porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	952
284	1	3	170	gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	84
284	2	282	1034	gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	712
284	3	1028	2026	gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	979
284	4	1990	2202	gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	187
289	3	1536	1991	gb M32470	S. aureus Sau3AI-restriction-enzyme and Sau3AI-modification-enzyme genes, complete cds	99	338
303	1	2	868	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	867
303	2	1409	2343	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	100	975
303	3	2367	3161	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	99	793
305	1	2707	1355	dbj D17366 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343
311	1	2628	1315	gb U42945	Staphylococcus aureus lysC and lycR genes, complete cds	98	1314
312	6	7019	7870	gb U44017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	74	351
323	1	1998	1003	gb U31175	Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996
326	1	1	237	emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	100	108
338	1	687	388	emb X64369 SALE	S. aureus leuF-P83 gene for F component of leucocidin R	98	259
338	2	1828	1088	emb X54389 SALE	S. aureus leuF-P83 gene for F component of leucocidin R	97	137

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	ORF nt length
342	2	579	1754	[gb U06462]	[Staphylococcus aureus SA4 FtsZ (ftsZ) gene, complete cds	100	1176
344	2	517	1248	[emb V01281 SNUU	[S. aureus mRNA for nuclease	98	732
349	1	457	230	[gb H20393]	[S. aureus bacteriophage phi-11 attachment site (attB)	96	172
353	1	1016	516	[gb HH3994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	187
353	2	1582	1046	[gb MB3994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	99	537
356	1	3	674	[gb U20503]	[Staphylococcus aureus MHC class II analog gene, complete cds	75	671
361	1	1	903	[gb L19298]	[Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	98	747
361	2	1103	1507	[gb L19298]	[Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	97	68
373	1	3	1148	[emb X62288 SAFE	[S. aureus DNA for penicillin-binding protein 2	99	1146
389	3	1904	1248	[emb X62282 SATS	[S. aureus target site DNA for IS411 insertion	97	349
400	1	1	540	[emb X61716 SNHL	[S. aureus hlb gene encoding sphingomyelinase	99	389
400	2	1693	1187	[emb X13404 SAHL	[Staphylococcus aureus hlb gene for beta-hemolysin	99	178
408	1	1810	1049	[gb S76213]	[asp23-alkaline shock protein 23 (methicillin resistant) (Staphylococcus aureus, 912, Genomic, 1360 nt)	99	163
418	1	2	217	[gb L41499]	[Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	216
418	2	854	639	[dbj D17366 STAA	[Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	100	168
421	2	1262	2509	[gb L43098]	[Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	1248
422	1	2	325	[gb K02985]	[S. aureus (strain RN450) transposon Tn554 insertion site	96	200
427	1	865	434	[dbj D28879 STAP	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	432
427	2	1829	1122	[dbj D28879 STAP	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	151
435	1	2	808	[dbj D86240 D862	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	556
435	2	832	959	[dbj D86240 D862	[Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	134
436	1	1341	665	[emb X17688 SAFE	[S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	97	657

TABLE I

TABLE I

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
436	2	2403	1657	emb N17688 SAFE	<i>S. aureus</i> factor essential for expression of methicillin resistance (<i>femA</i>) gene, complete cds, and <i>trpA</i> gene, 3' and	100	294	747
442	1	347	1300	emb X72700 SAPV	<i>S. aureus</i> genes for S and F components of Pantone-Valentine leucocidins	84	204	954
445	2	1906	2178	gb L01055	<i>Staphylococcus aureus</i> gamma-hemolysin components A, B and C (<i>hlgA</i> , <i>hlgB</i> , <i>hlgC</i>) genes, complete cds	98	187	273
447	1	167	1078	gb U19770	<i>Staphylococcus aureus</i> pyrrolidone carboxyl peptidase (<i>pcp</i>) gene, complete cds	100	51	912
447	2	1176	1784	gb U19770	<i>Staphylococcus aureus</i> pyrrolidone carboxyl peptidase (<i>pcp</i>) gene, complete cds	96	597	609
454	3	7309	4319	emb Z19852 SACF	<i>S. aureus</i> gene for clumping factor	75	653	2991
472	4	7896	5479	gb L25288	<i>Staphylococcus aureus</i> gyrase-like protein alpha and beta subunit (<i>grlA</i> and <i>grlB</i>) genes, complete cds	99	2418	2418
472	5	8120	6792	gb L25288	<i>Staphylococcus aureus</i> gyrase-like protein alpha and beta subunit (<i>grlA</i> and <i>grlB</i>) genes, complete cds	99	1328	1329
475	2	566	889	emb X52543 SAAG	<i>S. aureus</i> <i>agrA</i> , <i>agrB</i> and <i>hld</i> genes	100	76	324
481	4	1922	1560	emb X64172 SNRP	<i>S. aureus</i> <i>rplL</i> , <i>orf202</i> , <i>rpoB(rif)</i> and <i>rpoC</i> genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	250	363
481	5	1244	1534	emb X64172 SNRP	<i>S. aureus</i> <i>rplL</i> , <i>orf202</i> , <i>rpoB(rif)</i> and <i>rpoC</i> genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	224	291
487	2	1388	1188	gb H83994	<i>Staphylococcus aureus</i> prolipoprotein signal peptidase (<i>lsp</i>) gene, complete cds	98	72	201
489	1	2737	1370	gb U21221	<i>Staphylococcus aureus</i> hyaluronate lyase (<i>hyaA</i>) gene, complete cds	99	1368	1368
503	2	1135	653	gb H83994	<i>Staphylococcus aureus</i> prolipoprotein signal peptidase (<i>lsp</i>) gene, complete cds	100	108	483
511	3	1613	2242	gb L14017	<i>Staphylococcus aureus</i> methicillin-resistance protein (<i>mecR</i>) gene and unknown ORF, complete cds	84	323	630
511	4	3122	2700	gb S76213	<i>espD2</i> -alkaline shock protein 23 (methicillin resistant) [<i>Staphylococcus aureus</i> , 912, Genomic, 1360 nt]	96	423	423
520	2	758	1297	emb X72014 SAFI	<i>S. aureus</i> fib gene for fibrinogen-binding protein	99	540	540
520	3	1436	1801	emb X72013 SAFI	<i>S. aureus</i> fib gene for fibrinogen-binding protein	99	221	366
526	1	2150	1092	dbj D17366 STAA	<i>Staphylococcus aureus</i> <i>atl</i> gene for autolysin, complete cds and other ORFs	99	641	1059

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
528	2	58	963	[gb U19100]	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	260	906
528	3	1098	2870	[gb U19100]	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	866	1773
530	1	3	434	[gb U31979]	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	99	432	432
530	2	1211	2395	[gb U31979]	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	91	1185	1185
530	3	2409	2801	[gb U31979]	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	88	181	393
530	4	2690	3484	[gb U05004]	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	75	795
530	5	3482	4792	[gb U05004]	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	99	905	1311
530	6	4790	5380	[gb U05004]	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	196	591
539	1	3	338	[emb X76490 SAGL]	S. aureus (bb270) glnA and glnK genes	99	336	336
539	2	336	527	[emb X76490 SAGL]	S. aureus (bb270) glnA and glnK genes	100	189	192
534	1	727	365	[gb U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	54	363
534	2	2175	1252	[gb U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	918	924
534	3	1574	1374	[gb U73374]	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	96	122	201
584	2	1019	705	[gb U21221]	Staphylococcus aureus hyaluronate lyase (hlyA) gene, complete cds	99	306	315
587	3	1475	4288	[emb Z18852 SACF]	S. aureus gene for clumping factor	98	2588	2814
598	1	3881	1953	[dlj D28879 STAP]	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	99	1873	1929

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
605	1	2	745	[gb D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	98	338	744
609	1	1628	816	[emb X76490 SAGL	S. aureus (bb270) glnA and glnR genes	100	495	813
614	1	1280	642	[gb M32103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639
626	1	2508	1255	[gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	225	1251
626	2	3315	2284	[gb M63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	838	1032
629	1	1999	1001	[emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999
629	2	1407	1195	[emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213
631	2	5126	3228	[emb Z18852 SACP	S. aureus gene for clumping factor	82	489	1899
632	1	3	551	[emb Z30568 SAST	S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549
632	2	529	1323	[emb Z30568 SAST	S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795
651	1	1909	1070	[gb L19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	478	840
657	2	1800	1105	[gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696
662	1	908	456	[emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	369	453
662	2	230	475	[emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	246	246
662	3	746	1399	[emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	653	654
682	1	956	480	[gb M63177	S. aureus sigma factor (plac) gene, complete cds	100	136	477
685	1	1182	592	[gb U65000	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	98	534	591
685	2	1716	1153	[gb U65000	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	96	564	564
697	1	3	527	[gb M63177	S. aureus sigma factor (plac) gene, complete cds	100	195	525
697	2	485	784	[gb M63177	S. aureus sigma factor (plac) gene, complete cds	97	280	300

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
710	1	15	503	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	217	489
733	1	26	205	gb H80252	Staphylococcus aureus norA119 gene (which mediates active efflux of fluorquinolones), complete cds	97	140	180
741	1	1736	1197	dbj D81951 STAL	Staphylococcus aureus DNA for LukM component, LukP-PV like component, complete cds	81	522	540
752	1	1	636	emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	99	618	636
752	2	588	956	emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	99	340	369
756	1	1308	709	emb X01645 SATO	Staphylococcus aureus (Wood 46) gene for alpha-toxin	98	567	600
777	1	1582	950	emb Z49245 SA42	S. aureus partial sod gene for superoxide dismutase	99	429	633
780	1	1111	557	gb U20503	Staphylococcus aureus MHC class-II analog gene, complete cds	86	550	555
784	1	73	687	gb U63529	Staphylococcus aureus novel antigen gene, complete cds	99	568	615
797	1	182	544	dbj D14711 STAN	Staphylococcus aureus HSP10 and HSP60 genes	98	363	363
798	1	532	302	emb X58434 SAPD	S. aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	196	231
823	1	3	467	gb S77055	recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]	99	156	465
848	1	348	175	gb U25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and grlb) genes, complete cds	99	174	174
848	2	476	318	gb U25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and grlb) genes, complete cds	100	131	159
866	1	792	397	emb X64172 SNRP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 4 beta' chains	99	395	396
883	1	1	285	dbj D90119 STAN	S. aureus nraA gene	99	131	285
884	1	606	334	emb X52543 SAG	S. aureus agrA, agrB and hld genes	98	265	273
884	2	716	522	emb X52543 SAG	S. aureus agrA, agrB and hld genes	100	195	195
912	2	517	681	emb Z30588 SAST	S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	163	165
917	1	2	265	gb H64724	S. aureus tagatose 6-phosphate isomerase gene, complete cds	99	247	264
917	2	238	396	gb H64724	S. aureus tagatose 6-phosphate isomerase gene, complete cds	95	147	159
918	1	2426	1215	emb X93205 SAPT	S. aureus ptmH and ptmI genes	99	1212	1212

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	USP nt length	ORF nt length
967	1	1	411	dbj U90119 STAM	S. aureus nra gene	97	395	411
991	1	672	137	emb X52543 SAG	S. aureus agrA, agrB and hid genes	99	336	336
1000	1	1117	845	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecA) gene and unknown ORF, complete cds	78	190	273
1001	1	498	265	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	234	234
1010	1	1	285	gb U21221	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	224	285
1046	1	656	330	emb X72700 SAPV	S. aureus genes for S and P components of Pantone-Valentine leucocidins	85	205	327
1060	1	480	286	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolypsoamide acetyltransferase and dihydrolypsoamide dehydrogenase	99	180	195
1073	1	1176	589	gb K02985	S. aureus (strain RN450) transposon Tn554 insertion site	100	131	588
1079	1	3	230	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	228	228
1079	2	218	484	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	267	267
1079	1	460	645	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	186	186
1092	1	289	146	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolypsoamide acetyltransferase and dihydrolypsoamide dehydrogenase	98	124	144
1143	1	1	243	gi H63177	S. aureus sigma factor (plaC) gene, complete cds	99	243	243
1157	1	2	136	emb Z48003 SNDW	S. aureus gene for DNA polymerase III	97	127	135
1189	1	720	363	gb S74031	norA-nraA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	99	360	360
1190	1	2	283	gb H21854	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	282	282
1190	2	1127	888	emb X52543 SAG	S. aureus agrA, agrB and hid genes	100	240	240
1225	1	2	163	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	97	124	162
1243	1	2	529	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	495	528
1244	1	1	210	gi S74031	norA-nraA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	100	210	210
1301	1	41	472	emb X76490 SAGL	S. aureus (bb270) glpA and glpR genes	99	299	432

TABLE I

5. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1315	1	18	326	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	277	309
1519	1	2	175	dbj D28879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	dbj D66240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	98	672	672
1797	1	644	324	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	321	321
1857	1	1	192	gb H90536	Staphylococcus aureus alpha-hemolysin gene, 3' end	98	192	192
1923	1	2	181	emb X17688 SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	100	180	180
1957	1	2	346	gb U60589	Staphylococcus aureus novel antigen gene, complete cds	99	345	345
1988	1	1	402	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	402	402
2100	1	414	208	gb H63177	S.aureus sigma factor (plac) gene, complete cds	99	207	207
2199	1	1	402	gb U66664	Staphylococcus aureus DNA fragment with class II promoter activity	99	131	402
2537	1	308	156	emb X17688 SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	153	153
2891	1	2	400	gb J25426	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP70; HSP40; ORF35, complete cds	100	358	381
2971	1	3	398	gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mef), and o-succinylbenzoic acid synthetase (mnc) genes, complete cds	97	272	396
2978	1	618	328	gb J31979	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroascinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	98	250	291
2985	1	832	464	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	98	347	369
3006	1	2170	1784	gb U11779	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RBNV30 16S-23S rRNA spacer region	87	82	387
3008	1	474	238	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP70; HSP40; ORF35, complete cds	88	178	237
3008	2	451	281	dbj U30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP70; HSP40; ORF35, complete cds	97	120	171

TABLE 1

S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP length	ORF nt length
3011	1	793	198	emb X62992 SARF	S aureus fnbB gene for fibronectin binding protein B	93	72	396
3019	1	2	235	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234	234
3023	1	81	233	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	87	100	153
3029	1	90	287	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	100	135	198
3039	1	18	164	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	97	135	147
3039	2	70	327	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	77	183	258
3056	1	3	215	emb X64172 SARP	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213	213
3059	1	1	261	dbj J030690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	234	261
3073	1	27	284	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	258
3074	1	2	397	emb X64172 SARP	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250	396
3088	1	3	239	dbj D86727 D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	95	215	237
3097	1	444	244	emb Z48003 SARH	S aureus gene for DNA polymerase III	97	160	201
3102	1	307	155	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142	153
3121	1	568	398	emb X58434 SAPD	S aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	100	88	171
3125	1	463	233	emb A89233 SAMP	S aureus DNA for rpoC gene	98	192	231
3133	1	2	175	emb Z18852 SACF	S aureus gene for clumping factor	96	154	174
3160	1	420	211	dbj J10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	89	197	210
3176	1	1	378	emb X58434 SAPD	S aureus pdhA, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	91	378
3192	1	420	211	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72	210
3210	1	3	143	gb J176714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	96	141	141

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
3232	3	2106	1282	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	71	257
3538	1	2	394	emb X89233 SARP	S. aureus DNA for rpoC gene	99	350
3543	1	392	634	gb U11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	99	102
3555	1	637	320	emb Z18852 SACF	S. aureus gene for clumping factor	99	307
3559	1	3	182	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	100	141
3559	2	95	313	emb X17679 SALO	Staphylococcus aureus coa gene for coagulase	98	174
3563	1	278	141	gb U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	100	79
3563	2	527	363	gb U35773	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	98	162
3566	1	3	422	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	175
3588	1	2	262	gb U43098	Transposon TnS404 and insertion sequences IS181 and IS1182 (from Staphylococcus aureus) DNA	99	253
3593	1	3	350	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	345
3600	1	758	381	emb Z18852 SACF	S. aureus gene for clumping factor	72	346
3602	1	788	396	emb Z18852 SACF	S. aureus gene for clumping factor	98	319
3636	1	1013	528	emb Z18852 SACF	S. aureus gene for clumping factor	84	403
3682	1	3	236	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	231
3682	2	224	415	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	112
3693	1	758	423	emb X62992 SAFN	S. aureus fnbB gene for fibronectin binding protein B	100	229
3702	1	593	354	gb U11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	94	81
3725	1	924	463	emb Z18852 SACF	S. aureus gene for clumping factor	71	367
3761	1	809	450	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	95	333
3767	1	1	402	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	387

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
3775	1	2	286	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227	285
3786	1	456	229	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204	228
3786	2	542	366	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123	177
3798	1	3	251	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	99	249	249
3813	1	793	398	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	396	396
3819	1	184	402	emb X68425 SA23	S.aureus gene for 23S rRNA	99	161	219
3844	1	932	468	gb U8826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	87	204	465
3845	1	1	381	emb X58434 SAND	S.aureus pddH, pddC and pddD genes for pyruvate decarboxylase, dihydroliponate acetyltransferase and dihydroliponamide dehydrogenase	94	356	381
3856	1	798	400	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	76	192	399
3859	1	1049	573	emb Z18852 SACP	S.aureus gene for clumping factor	85	347	477
3871	1	650	327	gb N76716	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299	324
3876	1	2	253	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217	252
3877	1	572	288	gb J03479	S.aureus enzyme I ^{II} -lac (lacP), enzyme I ^I -lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	209	285
3878	1	1	237	emb X58434 SAND	S.aureus pddH, pddC and pddD genes for pyruvate decarboxylase, dihydroliponate acetyltransferase and dihydroliponamide dehydrogenase	96	155	237
3888	1	3	173	emb X16457 SA5T	Staphylococcus aureus gene for staphylocoagulase	98	171	171
3893	1	1	183	emb X89233 SARP	S.aureus DNA for rpoC gene	100	170	183
3893	2	181	357	emb X89233 SARP	S.aureus DNA for rpoC gene	98	79	177
3894	1	3	485	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450	483
3895	1	436	420	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	411	417
3905	1	48	239	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	159	192
3905	2	188	400	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	97	88	213

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3910	1	3	359	emb X58434 SAPP	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	278	357
3915	1	1	330	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	175	330
3964	1	691	347	emb Z48003 SADN	S. aureus gene for DNA polymerase III	100	295	345
4007	1	199	390	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	163	192
4036	1	3	371	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	339	369
4046	1	692	348	emb Z18852 SACF	S. aureus gene for clumping factor	87	221	345
4060	1	1	375	emb Z18852 SACF	S. aureus gene for clumping factor	96	271	375
4061	1	860	432	emb Z48003 SADN	S. aureus gene for DNA polymerase III	99	429	429
4062	1	606	304	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	198	303
4085	1	58	402	gb U11786	Staphylococcus aureus methicillin-resistant ATCC J3952 clone RMRV42 16S-23S rRNA spacer region	98	127	345
4088	1	2	301	gb F43098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	227	300
4093	1	2	277	emb X58434 SAPP	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	276	276
4097	1	1	402	emb Z18852 SACF	S. aureus gene for clumping factor	74	307	402
4116	1	22	402	gb L05004	Staphylococcus aureus dehydroquinase synthase (aroN) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (atoA) gene, complete cds; ORF3, complete cds	98	157	381
4125	1	240	401	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	86	162
4149	1	35	247	gb J04151	S. aureus fibronectin-binding protein (fmbA) mRNA, complete cds	99	200	213
4151	1	629	366	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	87	150	264
4154	1	754	398	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	297	357
4179	1	1	294	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	240	294

TABLE 1

S. aureus - Coding regions containing known sequences

Config ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	HSP nt length	ORF nt length
4203	1	1	235	emb X89233 SARP	S. aureus DNA for rpoC gene	99	239	255
4206	1	1	303	emb Z18852 SACF	S. aureus gene for clumping factor	100	236	303
4206	2	195	344	emb Z18852 SACF	S. aureus gene for clumping factor	95	65	150
4208	1	108	314	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	89	76	201
4216	1	656	330	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	326	327
4226	1	594	298	gb L11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	97	132	297
4260	1	216	383	gb U11784	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RN440 16S-23S rRNA spacer region	83	141	168
4272	1	355	179	emb Z48003 SAUN	S. aureus gene for DNA polymerase III	100	164	177
4276	1	4	177	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	99	150	174
4277	1	1	270	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	265	270
4282	1	691	377	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	282	315
4291	1	379	191	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	183	189
4295	1	3	329	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	94	144	327
4313	1	435	280	gb L11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	100	94	156
4315	1	3	185	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158	183
4315	2	101	310	gb J03479	S. aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75	210
4327	1	1	294	gb L143098	Transposon Tn3404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	294	294
4360	1	603	319	gb U02910	Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence	100	116	285
4364	1	3	146	emb X64172 SARP	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	95	140	146
4388	1	167	310	emb X62992 SAPN	S. aureus fnbB gene for fibronectin binding protein B	73	119	146

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
4401	1	2	313	emb X62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	97	243	312
4421	1	36	281	dbj D12572 STA2	Staphylococcus aureus rna gene for 23S ribosomal RNA	100	112	246
4426	1	3	293	emb Z18852 SACF	S.aureus gene for clumping factor	85	185	291
4428	1	493	248	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	139	246
4462	1	2	271	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	270	270
4466	1	1	240	emb Z18852 SACF	S.aureus gene for clumping factor	99	231	240
4469	1	1	312	gb J03479	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265	312
4485	1	3	263	gb L43098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	259	261
4492	1	74	400	gb M86227	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104	327
4497	1	515	269	emb Z18852 SACF	S.aureus gene for clumping factor	99	213	267
4529	1	2	172	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	151	171
4547	1	1	300	emb X62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	157	300
4554	1	318	160	emb Z18852 SACF	S.aureus gene for clumping factor	84	126	159
4565	1	9	227	emb Z18852 SACF	S.aureus gene for clumping factor	84	213	219
4569	1	79	223	emb Z18852 SACF	S.aureus gene for clumping factor	98	127	144
4608	1	22	216	emb X58434 SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	92	168	195
4614	1	464	234	emb Z18852 SACF	S.aureus gene for clumping factor	86	169	231
4623	1	105	302	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	152	198
4632	1	18	206	gb J03479	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183	189
4646	1	1	222	emb Z18852 SACF	S.aureus gene for clumping factor	84	100	222
4687	1	2	166	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	156	165

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
4695	1	313	158	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	155	156
4703	1	1	153	emb X5843 SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	103	153

TABLE I

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
20	6	5089	4679	gi 511819	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2032	1577	pir B49703 B497	int gene activator BlnA - bacteriophage phi 11	100	100	456
149	5	2109	1912	gi 166161	bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	100	100	198
349	2	558	409	gi 166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
398	1	1372	707	gi 166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	gi 455128	exciplonase (xle) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	1914	1744	gi 1204912	H. influenzae predicted coding region H10660 [Haemophilus influenzae]	100	71	171
849	1	2	262	gi 1373002	polyprotein [bean common mosaic virus]	100	46	261
1349	1	277	140	gi 143359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gi 49319	100	82	138
2880	1	21	308	gi 662933	protein kinase C inhibitor-1 [Homo sapiens]	100	98	288
3085	1	428	216	gi 1354211	PER112-like protein [Bacillus subtilis]	100	100	213
4168	2	571	398	gi 1354211	PER112-like protein [Bacillus subtilis]	100	100	174
331	1	2	247	gi 426473	nusG gene product [Staphylococcus carnosus]	98	95	246
207	2	1272	1463	gi 460259	enolase [Bacillus subtilis]	97	90	192
331	2	395	850	gi 381638	h11 protein [Staphylococcus carnosus]	97	93	456
366	1	39	215	gi 66161	bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	97	95	177
680	3	718	936	gi 426473	nusG gene product [Staphylococcus carnosus]	97	97	219
1578	1	284	144	gi 1319950	large subunit of NAMH-dependent glutamate synthase [Plectonura boryanum]	97	79	141
157	1	321	518	gi 1022726	unknown [Staphylococcus haemolyticus]	96	88	198
205	33	16470	16147	gi 1165302	S10 [Bacillus subtilis]	96	91	324
3919	1	48	401	gi 971784	Clp-like ATP-dependent protease binding subunit [Boa taurus]	96	81	354
4133	1	830	417	gi 1022726	unknown [Staphylococcus haemolyticus]	96	84	414
4168	1	708	355	gi 1354211	PER112-like protein [Bacillus subtilis]	96	95	354
4207	1	312	157	gi 402031	[similar to trimethylamine DH [Mycoplasma capricolum] pir S49950 S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC3) (fragment)]	96	86	156

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4227	2	152	331	gi 871784	Cip-like ATP-dependent protease binding subunit (Bos taurus)	96	81	180
4416	1	570	286	gi 1022726	unknown (Staphylococcus haemolyticus)	96	84	285
22	1	858	430	gi 511070	UreG (Staphylococcus xylosum)	95	88	429
22	7	4362	4036	gi 1581787	urease gamma subunit (Staphylococcus xylosum)	95	79	337
82	6	8794	9114	pir JC0008 JG00	ribosomal protein S7 - Bacillus stearothermophilus	95	81	321
154	9	9280	7838	gi 1354211	PGT12-like protein (Bacillus subtilis)	95	92	1443
186	3	2798	2055	gi 1514656	serine O-acetyltransferase (Staphylococcus xylosum)	95	87	744
205	5	4406	4014	gi 142462	ribosomal protein S11 (Bacillus subtilis)	95	85	393
205	7	5017	4793	gi 142459	initiation factor 1 (Bacillus subtilis)	95	84	225
205	21	11365	10991	gi 1044974	ribosomal protein L14 (Bacillus subtilis)	95	93	375
259	5	7288	6644	sp P47995 YSEA_	HYPOTHETICAL PROTEIN IN SECA 5'-REGION (ORF1) (FRAGMENT)	95	85	645
302	3	795	1097	gi 40186	homologous to E.coli ribosomal protein L27 (Bacillus subtilis) i 143592 L27 ribosomal protein (Bacillus subtilis) i C21895 C21895 ribosomal protein L27 - Bacillus subtilis p PU5457 R27_BACSU 50S RIBOSOMAL PROTEIN L27 (BLJ0) (BLJ4). i 40175 L28 gene prod	95	89	301
310	1	579	1523	gi 1177684	chorismate mutase (Staphylococcus xylosum)	95	92	945
414	1	2	163	pir C48396 C483	ribosomal protein L34 - Bacillus stearothermophilus	95	90	162
4185	2	125	277	gi 1276841	glutamate synthase (GOGAT) (Porphyra purpurea)	95	86	151
22	2	1028	723	gi 511069	UreF (Staphylococcus xylosum)	94	91	306
22	5	5066	3310	gi 410516	urease alpha subunit (Staphylococcus xylosum)	94	85	1737
60	4	815	1372	gi 666116	glucose kinase (Staphylococcus xylosum)	94	87	558
205	18	10012	9536	gi 1044978	ribosomal protein S8 (Bacillus subtilis)	94	78	477
326	4	3378	2542	gi 557492	dihydroxynaphthoic acid (DHNA) synthetase (Bacillus subtilis) gi 143186 dihydroxynaphthoic acid (DHNA) synthetase (Bacillus subtilis)	94	85	837
414	3	737	955	gi 467386	thiophen and furan oxidation (Bacillus subtilis)	94	77	219
426	3	2260	1823	gi 1263908	putative (Staphylococcus epidermidis)	94	87	438
514	1	2	355	gi 633650	enzyme I (mannitol) (Staphylococcus carnosus)	94	84	354
1017	1	2	229	gi 149435	putative (Lactococcus lactis)	94	73	228
3098	1	310	184	gi 411952	ipa-28d gene product (Bacillus subtilis)	94	50	147

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3232	1	630	316	gi 1022725	unknown [Staphylococcus haemolyticus]	94	84	315
42	5	2089	2259	pir B48396 B483	ribosomal protein L3 - Bacillus stearothermophilus	93	81	171
101	2	1745	1383	gi 155345	arsenic efflux pump protein [Plasmodium falciparum]	93	82	363
205	24	12227	11865	sp P14577 RL16_	50S RIBOSOMAL PROTEIN L16	93	83	363
259	4	8291	5673	gi 499335	secA protein [Staphylococcus carnosus]	93	85	2619
275	1	2226	1114	gi 631650	enzyme II(mannitol) [Staphylococcus carnosus]	93	86	1113
444	6	6207	5773	gi 1022726	unknown [Staphylococcus haemolyticus]	93	81	435
491	1	152	622	gi 46912	ribosomal protein L13 [Staphylococcus carnosus]	93	88	471
607	6	1674	2033	gi 1022726	unknown [Staphylococcus haemolyticus]	93	83	360
653	1	973	488	gi 580890	translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]	93	77	486
1864	1	3	194	gi 306553	ribosomal protein small subunit [Homo sapiens]	93	93	192
2997	1	28	300	gi 143390	carbamyl phosphate synthetase [Bacillus subtilis]	93	82	273
3332	2	907	596	gi 1022725	unknown [Staphylococcus haemolyticus]	93	84	312
3761	2	794	621	gi 1022725	unknown [Staphylococcus haemolyticus]	93	88	174
16	1	3	374	gi 142781	putative cytoplasmic protein; putative [Bacillus subtilis]	92	83	372
31	7	5915	6124	gi 1136430	sp P37954 UVRA_BACSU EXCINUCLEASE ABC SUBUNIT B (DNA PROTEIN) FRAGMENT	92	46	210
56	19	26483	27391	gi 467401	unknown [Bacillus subtilis]	92	80	909
69	6	5882	6130	gi 530200	trophoblastin [Ovis aries]	92	53	249
145	3	2568	2038	gi 1022725	unknown [Staphylococcus haemolyticus]	92	80	511
171	3	2760	2362	gi 517475	D-amino acid transaminase [Staphylococcus haemolyticus]	92	86	399
205	12	7495	6962	gi 49189	secY gene product [Staphylococcus carnosus]	92	85	534
205	19	10812	10255	gi 1044976	ribosomal protein L5 [Bacillus subtilis]	92	82	558
219	1	710	357	gi 1303812	Yqev [Bacillus subtilis]	92	88	354
344	3	1575	1805	gi 1405474	CspC protein [Bacillus cereus]	92	85	231
699	1	20	361	gi 413999	lpa-75d gene product [Bacillus subtilis]	92	81	342
1343	1	2	160	pir A45434 A454	ribosomal protein L19 - Bacillus stearothermophilus	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1958	1	524	264	gi 407908	Etfecr [Staphylococcus xylosum]	92	80	261
3578	2	718	386	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	92	78	333
3585	1	644	324	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	92	81	321
3640	1	4	402	gi 1022726	unknown [Staphylococcus haemolyticus]	92	81	399
4362	1	14	178	gi 450688	hcdM gene of Ecopri gene product [Escherichia coli] pir S18437 S18437 hcdM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	92	76	165
4446	1	358	182	gi 1022725	unknown [Staphylococcus haemolyticus]	92	82	177
4549	1	462	232	gi 1022726	unknown [Staphylococcus haemolyticus]	92	80	231
4626	1	3	224	gi 1022725	unknown [Staphylococcus haemolyticus]	92	84	222
2	4	1980	4531	gi 535349	CodW [Bacillus subtilis]	91	74	552
28	1	2	1126	gi 1001376	hypothetical protein [Synchocystis sp.]	91	78	1125
60	5	1354	1701	gi 1226043	orf2 downstream of glucose kinase [Staphylococcus xylosum]	91	80	348
101	1	1989	1036	gi 150728	arsenic efflux pump protein [Plasmid p1258]	91	80	954
187	2	412	1194	gi 142559	ATP synthase alpha subunit [Bacillus megaterium]	91	79	783
205	22	11579	11298	gi 40149	S17 protein (AA 1-87) [Bacillus subtilis]	91	83	282
206	7	8184	10262	gi 1072418	glcA gene product [Staphylococcus carnosus]	91	83	2079
306	2	3885	2326	gi 143012	GMP synthetase [Bacillus subtilis]	91	78	1560
306	3	5319	3826	gi 467399	IMP dehydrogenase [Bacillus subtilis]	91	79	1494
310	3	2194	3207	gi 1177685	ccpA gene product [Staphylococcus xylosum]	91	81	1014
343	4	2974	3150	gi 949974	sucrose repressor [Staphylococcus xylosum]	91	82	177
480	3	1606	3042	gi 433991	ATP synthase subunit beta [Bacillus subtilis]	91	85	1437
536	3	2026	1280	gi 143366	adenylosuccinate lyase [Pur-8] [Bacillus subtilis] pir C29326 W28SDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	91	79	747
552	1	1064	615	gi 297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir A49943 A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TH300)	91	79	450
637	1	1	1536	gi 143597	CTP synthetase [Bacillus subtilis]	91	79	1536
859	1	21	359	gi 385178	unknown [Bacillus subtilis]	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1327	1	339	530	gi1496558	orfX [Bacillus subtilis]	91	71	192
2515	1	466	275	gi1511070	ureC [Staphylococcus xylosus]	91	85	192
2594	1	2	202	gi1146824	beta-cyathionase [Escherichia coli]	91	75	201
3764	1	847	425	gi11022725	unknown [Staphylococcus haemolyticus]	91	78	423
4011	1	127	495	gi11022726	unknown [Staphylococcus haemolyticus]	91	75	369
4227	1	1	177	gi1296464	ATPase [Lactococcus lactis]	91	66	177
42	3	815	1033	gi1520401	catalase [Haemophilus influenzae]	90	86	219
51	8	3717	4607	gi1580899	OppF gene product [Bacillus subtilis]	90	74	891
129	3	5317	4001	gi1146206	glutamate dehydrogenase [Bacillus subtilis]	90	76	1317
164	17	16628	16933	sp P05766 RS15_	30S RIBOSOMAL PROTEIN S15 (BS18)	90	74	306
171	5	2983	2819	gi1517475	O-amino acid transaminase [Staphylococcus haemolyticus]	90	78	165
205	4	4497	3550	gi1142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	90	76	948
205	6	4748	4410	gi11044989	ribosomal protein S13 [Bacillus subtilis]	90	73	339
205	10	7165	6404	gi149189	secY gene product [Staphylococcus carnosus]	90	81	762
205	11	6645	6472	gi149189	secY gene product [Staphylococcus carnosus]	90	78	174
205	27	13602	13345	gi1786157	Ribosomal Protein S19 [Bacillus subtilis]	90	79	348
205	31	15858	15496	gi1165103	L3 [Bacillus subtilis]	90	79	363
260	5	7023	5773	gi1161380	icaA [Staphylococcus epidermidis]	90	78	1251
299	6	3378	3947	gi1467440	phosphoribosylpyrophosphate synthetase [Bacillus subtilis] gi140218 PRPP synthetase (AA 1-317) [Bacillus subtilis]	90	78	570
320	2	1025	1717	gi1312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	90	75	693
330	4	1581	1769	gi1986963	beta-tubulin [Sporidiobolus parvoseus]	90	80	189
369	1	954	523	gi1534762 S347	L-serine dehydratase beta chain - Clostridium sp.	90	77	432
557	1	3	188	gi1511589	M. Jannaschii predicted coding region MJ1624 [Methanococcus jannaschii]	90	54	186
663	2	667	1200	gi1143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pi1370481 WDS tryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus ubtillis	90	73	534
717	1	1	261	gi1143065	hubat [Bacillus stearothermophilus]	90	79	261
745	4	1059	865	gi1205433	H. influenzae predicted coding region H1190 [Haemophilus influenzae]	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1007	1	386	565	gi 143366	adenylosuccinate lyase (Pur-B) [Bacillus subtilis] pir C29326 W2B5DS	90	77	180
1054	1	579	331	gi 1033122	adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	90	50	249
1156	1	117	707	gi 1477776	ORF_729 [Escherichia coli]	90	80	591
1180	1	408	205	gi 1377831	ClpP [Bacillus subtilis]	90	74	204
1253	1	1	462	gi 40046	unknown [Bacillus subtilis]	90	75	462
2951	1	3	269	gi 144816	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]	90	76	267
3140	1	327	166	gi 1070014	ir S15936 NUSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	90	52	162
4594	1	3	233	gi 871784	formyltetrahydrofolate synthetase (FTHFS) (csg start codon) (EC 3.4.3)	90	76	231
87	1	1028	1750	gi 467327	protein-dependent [Bacillus subtilis]	89	75	723
112	1	2	505	gi 153741	Clp-like ATP-dependent protease binding subunit [Bos taurus]	89	77	504
118	1	120	398	gi 1303804	unknown [Bacillus subtilis]	89	75	279
128	4	3545	3757	gi 460257	ATP-binding protein [Streptococcus mutans]	89	84	213
164	12	11667	12755	gi 39954	YqeQ [Bacillus subtilis]	89	80	1089
205	13	7875	7405	gi 216338	triose phosphate isomerase [Bacillus subtilis]	89	76	471
205	32	16152	15823	gi 1165303	IF2 (aa 1-741) [Bacillus stearothermophilus]	89	80	330
270	3	2407	2207	pir C41902 C419	ORF for L15 ribosomal protein [Bacillus subtilis]	89	81	201
395	2	157	672	gi 520574	L3 [Bacillus subtilis]	89	80	516
494	1	3	839	gi 396259	arsenate reductase (EC 1.-.-.-) - Staphylococcus xylosus plasmid pSX267	89	77	837
510	1	1	444	gi 40046	glutamate racemase [Staphylococcus haemolyticus]	89	74	444
615	1	2124	1210	gi 1303812	protease [Staphylococcus epidermidis]	89	74	915
841	1	18	301	gi 1165303	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus]	89	80	324
1111	1	352	813	gi 47146	ir S15936 NUSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	89	70	462
1875	1	2	256	gi 1203108	thermonuclease [Staphylococcus intermedius]	89	82	255
2963	1	11	367	gi 467458	ATP-dependent protease binding subunit [Haemophilus influenzae]	89	83	357
					cell division protein [Bacillus subtilis]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3020	1	90	362	gi1239988	hypothetical protein [Bacillus subtilis]	89	66	273
3565	1	2	400	gi1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	89	75	399
3586	1	105	314	gi1580832	ATP synthase subunit gamma [Bacillus subtilis]	89	82	210
3629	1	794	399	gi1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	78	396
3688	1	2	400	gi1146206	glutamate dehydrogenase [Bacillus subtilis]	89	75	399
3699	1	794	399	gi1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	89	75	396
4016	1	428	216	gi1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	71	213
4177	1	471	301	gi1149426	putative [Lactococcus lactis]	89	76	171
4436	1	601	302	gi1022725	unknown [Staphylococcus haemolyticus]	89	80	300
4635	1	320	162	gi1022725	unknown [Staphylococcus haemolyticus]	89	73	159
2	2	1330	2676	gi1520754	putative [Bacillus subtilis]	88	76	1347
42	2	468	848	ep1423231 CATA_	CATALASE (EC 1.11.1.61)	88	76	381
53	5	6389	4722	gi1474177	alpha-D-1,4-glucosidase [Staphylococcus xylosus]	88	80	1668
56	16	18018	18617	gi1467411	recombination protein [Bacillus subtilis]	88	77	600
60	3	376	843	gi1666116	glucose kinase [Staphylococcus xylosus]	88	77	468
70	2	1583	1245	gi144095	replication initiator protein [Listeria monocytogenes]	88	74	339
82	8	11514	12719	gi1460663 A606	translation elongation factor Tu - Bacillus subtilis	88	79	1206
103	7	4179	4391	gi1167181	serine/threonine kinase receptor [Brassica napus]	88	77	213
114	8	7732	8232	gi1022726	unknown [Staphylococcus haemolyticus]	88	72	501
118	2	308	2011	gi1303804	YqeQ [Bacillus subtilis]	88	77	1704
141	3	657	1136	gi1405446	transketolase [Bacillus subtilis]	88	72	480
148	7	5873	6116	gi1118002	dihydropterate synthase [Staphylococcus haemolyticus]	88	78	246
165	3	1428	2231	gi140053	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis]	88	80	804
205	28	15027	14185	gi1165306	trpS1730 yfpSA phenylalanine--tRNA ligase (EC 6.1.1.20) alpha ain - Bacillus subtilis			
225	1	1569	898	gi1303840	L2 [Bacillus subtilis]	88	82	843
235	1	2	1975	gi1452309	YqfS [Bacillus subtilis]	88	78	672
					valyl-tRNA synthetase [Bacillus subtilis]	88	76	1974

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
339	3	2060	1566	gi1118002	[dihydropterate synthase (Staphylococcus haemolyticus)]	88	73	495
443	4	4325	2928	gi1558559	[pyrimidine nucleoside phosphorylase (Bacillus subtilis)]	88	73	1398
532	1	3	419	gi1143797	[valyl-tRNA synthetase (Bacillus stearothermophilus) sp P11931 SYV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE) (VALRS)]	88	78	417
534	3	2504	2968	gi153049	[mannitol-specific enzyme-III (Staphylococcus carnosus) pir JQ0088 JQ0088 phosphotransferase system enzyme II (EC 7.1.1.69), mannitol-specific, factor III - Staphylococcus carnosus sp P17876 PTNA_STACA PTS SYSTEM, MANNITOL-SPECIFIC IIA COMPONENT E1A-NTL)]	88	82	465
705	2	584	399	gi1710018	[nitrite reductase (nirB) (Bacillus subtilis)]	88	70	186
1000	2	1824	1309	gi1022726	[unknown (Staphylococcus haemolyticus)]	88	78	516
1299	1	587	324	gi1401786	[phosphoenolpyruvate decarboxylase (Mycobacterium tuberculosis)]	88	55	264
1341	2	170	400	gi139963	[ribosomal protein L20 (AA 1-119) (Bacillus stearothermophilus) ir S05346 K5B520 ribosomal protein L20 - Bacillus stearothermophilus]	88	82	231
1386	1	41	214	gi1847154 B471	[signal recognition particle 54K chain homolog Ffh - Bacillus subtilis]	88	71	174
1386	2	183	533	gi1847154 B471	[signal recognition particle 54K chain homolog Ffh - Bacillus subtilis]	88	73	351
2949	1	704	399	gi1535350	[CodX (Bacillus subtilis)]	88	73	306
2984	1	5	169	gi1218277	[O-acetylserine(thiol) lyase (Spinacia oleracea)]	88	70	165
3035	1	1	138	gi1493083	[dihydroxyacetone kinase (Citrobacter freundii)]	88	67	138
3089	1	3	152	gi1606055	[ORF_F746 (Escherichia coli)]	88	88	150
3917	1	817	410	gi1143378	[pyruvate decarboxylase (E-1) beta subunit (Bacillus subtilis) gi1377836 pyruvate decarboxylase E-1 beta subunit (Bacillus ubtilis)]	88	77	408
4199	1	680	342	gi11405454	[aconitase (Bacillus subtilis)]	88	82	339
4201	1	734	369	gi1515938	[glutamate synthase (ferredoxin) (Synecocystis sp.) pir S46957 S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.]	88	84	366
4274	1	1	336	gi1515938	[glutamate synthase (ferredoxin) (Synecocystis sp.) pir S46957 S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.]	88	84	336
4308	1	794	399	gi1146206	[glutamate dehydrogenase (Bacillus subtilis)]	88	71	396
2	5	4570	6000	gi1535350	[CodX (Bacillus subtilis)]	87	70	1431
52	8	6781	6482	gi11064791	[function unknown (Bacillus subtilis)]	87	66	300

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
73	3	1584	2480	gi 142992	glycerol kinase (gpk) (EC 2.7.1.30) [Bacillus subtilis] pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL - PHOSPHOTRANSFERASE) (GLYCEROKINASE) (OR)	87	72	897
98	12	8813	9100	gi 467433	unknown [Bacillus subtilis]	87	62	288
124	4	4265	2988	gi 556886	serine hydroxymethyltransferase [Bacillus subtilis] pir S49363 S49363 serine hydroxymethyltransferase - Bacillus ubtilla	87	77	1278
124	6	4457	4032	gi 556883	Unknown [Bacillus subtilis]	87	66	426
148	5	3741	4559	gi 467460	unknown [Bacillus subtilis]	87	70	819
164	13	12710	13810	gi 39954	IF2 (aa 1-741) [Bacillus stearothermophilus]	87	72	1101
177	2	1104	2126	gi 467385	unknown [Bacillus subtilis]	87	78	1023
199	1	1982	1158	gi 141527	iron-sulfur protein [Bacillus subtilis]	87	77	825
199	2	4717	2933	pir A27763 A277	succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis	87	80	1785
205	23	11782	11543	gi 1044972	ribosomal protein L29 [Bacillus subtilis]	87	78	240
205	25	13275	13607	gi 1165309	S3 [Bacillus subtilis]	87	75	669
222	1	2033	1107	gi 1177249	rec23 gene product [Bacillus subtilis]	87	70	927
236	3	1635	1333	gi 1146198	ferredoxin [Bacillus subtilis]	87	80	303
246	5	2585	2292	gi 467373	ribosomal protein S18 [Bacillus subtilis]	87	77	294
260	2	4189	3422	gi 1161382	IcaC [Staphylococcus epidermidis]	87	72	768
320	3	1696	2391	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolycicus]	87	80	696
380	4	1165	1383	gi 142570	ATP synthase c subunit [Bacillus firmus]	87	80	219
414	4	900	1073	gi 467386	thiophen and furan oxidation [Bacillus subtilis]	87	77	174
435	2	1003	794	gi 1046166	pilin repressor [Mycoplasma genitalium]	87	69	210
448	1	1255	722	gi 405134	acetate kinase [Bacillus subtilis]	87	75	534
480	1	1	711	gi 142559	ATP synthase alpha subunit [Bacillus megaterium]	87	79	711
481	1	2	352	sp Q06797 RL1.8	50S RIBOSOMAL PROTEIN L1 (BL1)	87	72	351
677	2	359	955	gi 460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	597
677	3	934	1284	gi 460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	351
876	1	3	452	gi 1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	87	79	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Cuntig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1376	1	426	214	gi11065555	[F46H6.4 gene product (Caenorhabditis elegans)]	87	75	213
2206	1	3	376	gi1215098	[excisionase (Bacteriophage 154a)]	87	72	372
2938	1	3	290	gi1508979	[GTP-binding protein (Bacillus subtilis)]	87	69	288
3081	2	126	308	gi1467399	[IMP dehydrogenase (Bacillus subtilis)]	87	72	183
3535	1	3	401	gi11405454	[aconitase (Bacillus subtilis)]	87	80	399
4238	1	547	275	gi1603769	[HutU protein, urocanase (Bacillus subtilis)]	87	73	273
4	8	10427	8736	gi1603769	[HutU protein, urocanase (Bacillus subtilis)]	86	72	1692
22	6	4190	3738	gi1410515	[urease beta subunit (Staphylococcus xylosum)]	86	73	453
54	2	2480	1572	gi1289287	[UDP-glucose pyrophosphorylase (Bacillus subtilis)]	86	70	909
124	3	2336	1713	gi1556887	[uracil phosphoribosyltransferase (Bacillus subtilis)]	86	74	624
				gi1556887	[uracil phosphoribosyltransferase - Bacillus subtilis]			
148	3	1349	3448	gi1467458	[cell division protein (Bacillus subtilis)]	86	75	2100
148	4	3638	3859	gi1467460	[unknown (Bacillus subtilis)]	86	73	222
152	3	1340	2086	gi1377835	[pyruvate decarboxylase E-1 alpha subunit (Bacillus subtilis)]	86	75	747
164	18	17347	19467	gi1184680	[polynucleotide phosphorylase (Bacillus subtilis)]	86	72	2121
180	2	554	1159	gi1143467	[ribosomal protein S4 (Bacillus subtilis)]	86	80	606
205	3	2966	2592	gi1142464	[ribosomal protein L17 (Bacillus subtilis)]	86	77	375
205	26	13364	12990	gi140107	[ribosomal protein L22 - Bacillus earothermophilus]	86	75	375
246	7	3463	3140	gi1467375	[ribosomal protein S6 (Bacillus subtilis)]	86	70	324
299	3	1196	1540	gi139656	[spoVG gene product (Bacillus megaterium)]	86	70	345
299	7	3884	4345	gi1467440	[phosphoribosylpyrophosphate synthetase (Bacillus subtilis)]	86	78	462
					[synthetase (AA 1-317) (Bacillus subtilis)]			
304	5	2170	2523	gi1666983	[putative ATP binding subunit (Bacillus subtilis)]	86	65	354
310	2	1487	1678	gi11177684	[chorismate mutase (Staphylococcus xylosum)]	86	71	192
337	5	2086	3405	gi1487434	[isocitrate dehydrogenase (Bacillus subtilis)]	86	78	1320
339	2	1489	1109	gi1118003	[dihydropyrimidin aldolase (Staphylococcus haemolyticus)]	86	77	381
358	2	2124	3440	gi1146219	[28.3% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)]	86	73	1317

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% idenc	length (nt)
404	2	1015	2058	gi1303817	YnfA [Bacillus subtilis]	86	78	1044
581	2	661	452	gi140056	phoP gene product [Bacillus subtilis]	86	71	210
642	2	138	1075	gi1176399	EpIF [Staphylococcus epidermidis]	86	72	738
770	1	622	347	gi1143328	phoP protein [put.], putative [Bacillus subtilis]	86	69	276
865	1	1777	890	gi11146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	85	74	888
868	2	963	1133	gi11002911	transmembrane protein [Saccharomyces cerevisiae]	86	59	171
904	1	1	162	gi11303912	YghW [Bacillus subtilis]	86	72	162
989	1	35	433	gi11303993	YqkL [Bacillus subtilis]	86	76	399
1212	1	296	150	gi1414014	ipa-90d gene product [Bacillus subtilis]	86	70	147
1323	1	2	148	gi140041	pyruvate dehydrogenase (lipoamide) [Bacillus stearothermophilus] ir S10798 DESPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus	86	75	147
3085	2	540	310	gi11354211	PER12-like protein [Bacillus subtilis]	86	86	231
1847	1	1	228	nl 296464	ATPase [Lactococcus lactis]	86	63	228
4487	1	476	240	gi11022726	unknown [Staphylococcus haemolyticus]	86	73	237
4583	1	372	187	gi11022725	unknown [Staphylococcus haemolyticus]	86	79	186
25	5	4287	5039	gi11502421	3-ketoacyl-acyl carrier protein reductase [Bacillus subtilis]	85	64	753
56	21	30627	29395	gi11408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	85	69	1233
68	2	332	1192	gi1467376	unknown [Bacillus subtilis]	85	74	861
73	2	880	1707	gi1142992	glycerol kinase (gpk) (EC 2.7.1.30) [Bacillus subtilis] pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis ap B18157 GUPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	85	72	828
106	4	1505	3490	gi1143766	(thrSv) (EC 6.1.1.3) [Bacillus subtilis]	85	74	1986
128	2	1153	2202	gi1311924	glyceraldhyde-3-phosphate dehydrogenase [Clostridium pasteurianum] pir S34254 S34254 glyceraldehyde-3-phosphate dehydrogenase (EC 2.1.1.12) - Clostridium pasteurianum	85	75	1050
129	4	6466	5257	gi11064807	ORTHOMINE AMINOTRANSFERASE [Bacillus subtilis]	85	73	1215
138	6	3475	5673	gi11072419	glcB gene product [Staphylococcus carnosus]	85	74	2199
189	1	2	169	gi1467385	unknown [Bacillus subtilis]	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
205	15	8624	8106	gi 1044981	ribosomal protein S5 [Bacillus subtilis]	85	75	519
205	20	10928	10596	gi A02819 RSBS	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi 48980	secA gene product [Bacillus subtilis]	85	66	390
231	4	4877	3159	gi 1002520	MutS [Bacillus subtilis]	85	70	1719
243	9	8013	8783	gi 414011	ipa-87r gene product [Bacillus subtilis]	85	72	771
249	2	5894	3186	gi 1405454	aconitase [Bacillus subtilis]	85	73	2709
302	1	140	475	gi 40173	homolog of E. coli ribosomal protein L21 [Bacillus subtilis] ir S18439 S18439 Ribosomal protein L21 - Bacillus subtilis P126908 RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20)	85	72	336
333	1	5445	2968	gi 442360	CipC adenosine triphosphatase [Bacillus subtilis]	85	69	2478
364	6	6082	8196	gi 871784	Cip-like ATP-dependent protease binding subunit [Bos taurus]	85	68	2115
448	2	1992	1339	gi 405334	acetate kinase [Bacillus subtilis]	85	68	654
747	1	1251	853	gi 1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	85	73	399
886	2	159	467	gi 541768	hemin permease (Versinia enterocolitica)	85	55	309
1089	1	1208	606	pir B47154 B471	signal recognition particle SVK chain homolog Ffh - Bacillus subtilis	85	71	603
1163	1	416	409	gi 304155	diaminopimelate decarboxylase [Bacillus methanolicus] sp P41023 DCDA_BACHT DIAMINOPIMELATE DECARBOXYLASE [EC 4.1.1.20] DAP DECARBOXYLASE	85	62	408
1924	1	487	251	gi 215098	excisionase [Bacteriophage 154a]	85	73	237
2932	1	776	390	gi 1041099	Pyruvate Kinase [Bacillus licheniformis]	85	71	387
3030	1	3	275	gi 42370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir S01788 S01788 formate C-acetyltransferase [EC 2.3.1.54] - chierichia coli	85	74	273
3111	1	595	299	gi 63568	limb deformity protein [Gallus gallus]	85	85	297
3778	1	630	316	gi 391840	beta-subunit of Hdr [Pseudomonas fragi]	85	67	315
3835	1	1	387	gi 1204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	387
4042	1	3	386	gi 18178	formate acetyltransferase [Chlamydomonas reinhardtii] ir S24997 S24997 formate C-acetyltransferase [EC 2.3.1.54] - lamydomonas reinhardtii	85	70	384
4053	1	35	340	gi 1204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	306
4108	1	2	181	gi 1072418	glcA gene product [Staphylococcus carnosus]	85	61	180
4300	1	575	330	gi 151932	fructose enzyme II [Rhodobacter capsulatus]	85	59	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4392	1	627	355	gi11022725	unknown [Staphylococcus haemolyticus]	85	74	273
4408	1	2	235	gi1871784	clp-like ATP-dependent protease binding subunit [Bos taurus]	85	62	234
4830	1	578	291	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	85	68	288
4555	1	2	253	gi1450688	hadM gene of Ecopri gene product [Escherichia coli] pir[S38437/S38437 hadM protein - Escherichia coli pir[S09629/S09629 hypothetical protein A - Escherichia coli (SUB 40-520)]	85	52	252
4611	1	481	242	gi1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	85	65	240
4	10	1061	10591	gi146982	fosB gene product [Staphylococcus epidermidis]	84	68	511
13	2	1348	1172	gi1142450	ahcC protein [Bacillus subtilis]	84	56	177
16	4	1803	4652	gi1277198	DNA repair protein [Deinococcus radiodurans]	84	67	2850
22	3	1535	1128	gi1511069	lureP [Staphylococcus xylosum]	84	73	408
23	7	5055	5306	gi1603320	Yer082p [Saccharomyces cerevisiae]	84	61	252
53	11	11597	11145	gi1303948	YqjW [Bacillus subtilis]	84	68	453
53	12	14059	12770	gi1142613	branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis]	84	71	1290
70	1	1332	982	gi146647	ORF (repE) [Staphylococcus aureus]	84	68	351
73	4	2512	4331	gi1142993	glycerol-3-phosphate dehydrogenase (gldP) (EC 1.1.99.5) [Bacillus subtilis]	84	74	1800
98	7	4324	6096	gi1467427	methionyl-tRNA synthetase [Bacillus subtilis]	84	66	1773
100	9	9501	8680	gi1340128	ORF1 [Staphylococcus aureus]	84	78	822
117	3	1934	3208	gi1237019	Srb [Bacillus subtilis]	84	68	1275
148	6	4720	5670	gi1467462	cysteine synthetase A [Bacillus subtilis]	84	69	951
152	4	2064	2456	gi1143377	pyruvate decarboxylase (E-1) alpha subunit [Bacillus subtilis]	84	70	393
169	7	3634	3861	gi11001342	hypothetical protein [Synecocystis sp.]	84	66	228
171	4	2992	2657	gi1517475	D-alanine acid transaminase [Staphylococcus haemolyticus]	84	71	336
186	6	6941	6216	gi1467475	unknown [Bacillus subtilis]	84	70	726
205	9	6261	5692	gi1216340	ORF for adenylate kinase [Bacillus subtilis]	84	71	570
224	2	915	1391	gi1288269	beta-fructofuranosidase [Staphylococcus xylosum]	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match (%)	Accession	Match (%)	Gene name	% sim	% ident	length (nt)
251	1	92	388	91	gi1303790	91	YqeI [Bacillus subtilis]	84	65	297
282	3	1526	2836	91	gi1143040	91	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pic[D42728]D42728 glutamate-1-semialdehyde 2,1-aminomutase [EC 4.3.8] - Bacillus subtilis	84	75	1311
307	5	3138	2959	91	gi1070014	91	protein-dependent [Bacillus subtilis]	84	62	180
320	4	2343	4229	91	gi1143390	91	carbamyl phosphate synthetase [Bacillus subtilis]	84	70	1887
372	1	3	296	91	gi1022725	91	unknown [Staphylococcus haemolyticus]	84	70	294
413	2	2201	1341	91	gi11256146	91	YbbQ [Bacillus subtilis]	84	65	861
439	1	3	392	91	gi1046173	91	osmotically inducible protein [Mycoplasma genitalium]	84	53	390
461	3	1362	2270	91	gi140211	91	threonine synthase [thrC] (AA 1-352) (Bacillus subtilis) ir[A25364]A25364 threonine synthase [EC 4.2.99.2] - Bacillus btills	84	69	909
487	1	3	299	91	gi1114531	91	Integrin-like protein alpha Intlp [Candida albicans]	84	46	297
491	2	624	905	91	gi1508564 A305	91	ribosomal protein S9 - Bacillus stearothermophilus	84	69	282
491	3	816	1033	91	gi1508564 A305	91	ribosomal protein S9 - Bacillus stearothermophilus	84	77	198
548	1	3	341	91	gi1431231	91	Unacil permaso [Bacillus caldolyticus]	84	74	339
728	2	2701	1748	91	gi1912445	91	DNA polymerase [Bacillus caldolyticus]	84	68	954
769	1	3	257	91	gi1510953	91	Icobalamin biosynthesis protein N [Methanococcus jannaschii]	84	38	255
954	1	308	156	91	gi1405454	91	Iaconitase [Bacillus subtilis]	84	57	151
957	1	3	395	91	gi143402	91	recombination protein (ttg start codon) [Bacillus subtilis] gi1303923 RecW [Bacillus subtilis]	84	68	393
975	1	3	452	91	gi1885934	91	CipB [Synecococcus sp.]	84	70	450
1585	1	3	257	91	gi1510140	91	Ilgendopeptidase F [Lactococcus lactis]	84	56	255
2954	1	3	323	91	gi1603769	91	HutU protein, urocanase [Bacillus subtilis]	84	73	321
2996	1	650	348	91	gi181178	91	formate acetyltransferase [Chlamydomonas reinhardtii] ir[S24997]S24997 formate C-acetyltransferase [EC 2.3.1.54] - lamydomonas reinhardtii	84	65	303
3766	1	737	375	91	gi1517205	91	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	84	72	363
4022	1	2	169	91	gi11146206	91	glutamate dehydrogenase [Bacillus subtilis]	84	54	168
4058	1	620	312	91	gi151932	91	fructose enzyme 11 [Rhodospirillum rubrum]	84	71	309
4108	2	106	351	91	gi1072418	91	glcA gene product [Staphylococcus carnosus]	84	77	246

TABLE 2

S aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4183	1	3	308	gi 603769	lutU protein, urocanase [Bacillus subtilis]	84	72	306
4726	1	55	234	gi 146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pir A29617/A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large chain - Escherichia coli	84	73	180
22	4	2043	1576	gi 1393297	urease accessory protein [Bacillus sp.]	83	64	468
53	13	14722	13745	gi 142612	branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]	83	68	978
57	16	13357	12872	gi 143132	lactate dehydrogenase (AC 1.1.1.27) [Bacillus caldolyticus] pir B29704/B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus caldolyticus	83	66	486
66	3	3119	2274	gi 1303894	Yqhm [Bacillus subtilis]	83	63	886
66	5	6118	4643	gi 1212730	Yqhk [Bacillus subtilis]	83	68	1476
70	3	1864	1523	gi 144095	replication initiator protein [Listeria monocytogenes]	83	73	342
90	1	377	1429	gi 155571	alcohol dehydrogenase 1 (adhA) (EC 1.1.1.1) [Zymomonas mobilis] pir A35260/A35260 alcohol dehydrogenase (EC 1.1.1.1) - Zymomonas mobilis	83	70	1053
95	2	708	2162	gi 506381	phospho-beta-glucosidase [Bacillus subtilis]	83	70	1455
137	1	68	694	gi 467391	initiation protein of replication [Bacillus subtilis]	83	77	627
140	4	3209	2742	gi 634107	ldpb [Escherichia coli]	83	65	468
142	3	3468	2989	gi 1212776	lumasine synthase (b-subunit) [Bacillus amyloliquefaciens]	83	69	480
161	12	5749	6696	gi 903307	ORF75 [Bacillus subtilis]	83	64	948
164	9	9880	11070	gi 49316	ORF2 gene product [Bacillus subtilis]	83	66	1191
164	14	14148	14546	gi 580902	ORF6 gene product [Bacillus subtilis]	83	60	399
170	2	3144	2467	gi 520844	orf4 [Bacillus subtilis]	83	64	678
186	2	2029	1370	gi 289284	cysteineyl-tRNA synthetase [Bacillus subtilis]	83	72	660
205	14	7822	7607	gi 216337	ORP for L30 ribosomal protein [Bacillus subtilis]	83	74	216
217	6	3693	4540	gi 1510488	imidazoleglycerol-phosphate synthase (cyclase) [Methanococcus jannaschii]	83	60	858
301	1	985	638	gi 467419	unknown [Bacillus subtilis]	83	65	348
302	4	1421	2743	gi 508979	GTP-binding protein [Bacillus subtilis]	83	68	1323
321	4	3933	3571	gi 39844	fumarate (cicG) (aa 1-462) [Bacillus subtilis]	83	68	363
367	1	2	352	gi 1039479	ORFU [Lactococcus lactis]	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
387	1	3	662	gi 806281	DNA polymerase I [Bacillus stearothermophilus]	83	70	660
527	2	916	1566	gi 396259	protease [Staphylococcus epidermidis]	83	67	651
533	1	355	179	gi 142455	alanine dehydrogenase [EC 1.4.1.1] [Bacillus stearothermophilus] pir B34261 B34261 alanine dehydrogenase [EC 1.4.1.1] - Bacillus stearothermophilus	83	66	177
536	4	1617	1438	gi 143366	adenylosuccinate lyase (Pur-B) [Bacillus subtilis] ; tr C29336 W28SDS adenylosuccinate lyase [EC 4.3.2.2] - Bacillus subtilis	83	67	180
652	1	2	859	gi 520753	DNA topoisomerase I [Bacillus subtilis]	83	72	858
774	2	200	361	gi 1522665	M. jannaschii predicted coding region MJECU28 [Methanococcus jannaschii]	83	58	162
897	1	120	296	gi 1064807	ORTHINE AMINOTRANSFERASE [Bacillus subtilis]	83	76	177
1213	1	3	491	gi 289288	lexA [Bacillus subtilis]	83	67	489
2529	1	296	150	gi 143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir JY0481 YMB5 tryptophan-tRNA ligase [EC 6.1.1.2] - Bacillus subtilis	83	69	147
2973	1	649	326	gi 1109687	ProZ [Bacillus subtilis]	83	58	324
1009	1	728	366	gi 1482532	ORF_0394 [Escherichia coli]	83	65	363
3035	2	45	305	gi 950062	hypothetical yeast protein 1 [Mycoplasma capricolum] pir S48578 S48578 hypothetical protein - Mycoplasma capricolum SGC3 (fragment)	83	59	281
3906	1	67	309	gi 1353197	thioredoxin reductase [Bubacterium acidaminophilum]	83	61	243
4458	1	540	271	gi 197526	clumping factor [Staphylococcus aureus]	83	78	270
4570	1	444	223	gi 1022726	unknown [Staphylococcus haemolyticus]	83	74	222
4654	1	97	261	gi 1072419	glcB gene product [Staphylococcus carnosus]	83	79	165
16	2	295	1191	gi 153854	jvs402 protein [Streptococcus pneumoniae]	82	67	897
16	3	1193	1798	gi 153854	jvs402 protein [Streptococcus pneumoniae]	82	70	606
38	12	9644	8724	gi 1204400	N-acetylneuraminate lyase [Haemophilus influenzae]	82	58	921
42	4	988	2019	gi 841192	catalase [Bacteroides fragilis]	82	70	1032
51	6	2590	3489	gi 143607	sporulation protein [Bacillus subtilis]	82	69	900
56	11	12270	13925	gi 39431	oligo-1,6-glucosidase [Bacillus cereus]	82	60	1656
56	15	17673	18014	gi 467410	unknown [Bacillus subtilis]	82	66	342
61	2	881	3313	gi 143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	70	2433

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
82	7	9162	11318	gi148240	elongation factor G (AA 1-691) [Thermus aquaticus thermophilus] lr S15928 EFTMG translation elongation factor G - Thermus aquaticus p P13551 EFG_THERM ELONGATION FACTOR G (EFG-G)	82	64	2157
85	2	5470	3260	gi1143369	phosphoribosylformyl glycylamide synthetase II (PUR-Q) [Bacillus subtilis]	82	66	2211
102	6	3662	5380	gi1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	82	65	1719
117	4	3282	3493	pir A47154 A471	orf1 5' of Pth - Bacillus subtilis	82	53	252
128	6	4377	5933	gi1460258	phosphoglycerate mutase [Bacillus subtilis]	82	66	1557
129	2	1229	2182	gi1403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir S37251 S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	82	62	954
170	1	2	1441	gi1377831	unknown [Bacillus subtilis]	82	67	1440
177	1	3	1094	gi1467386	thiophen and furan oxidation [Bacillus subtilis]	82	65	1092
184	4	3572	4039	gi1153566	ORP (19K protein) [Enterococcus faecalis]	82	59	468
189	8	4453	4225	gi11001878	CspL protein [Listeria monocytogenes]	82	73	231
206	19	21366	20707	gi1473916	lipopeptide antibiotics Iturin A [Bacillus subtilis] sp P39144 LPI4_BACSU LIPPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN BIOSYNTHESIS PROTEIN.	82	50	660
221	2	805	1722	gi1517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	82	63	918
223	4	3866	3651	gi1439619	[Salmonella typhimurium IS200 insertion sequence from SAR17, attial.], gene product [Salmonella typhimurium]	82	69	216
260	3	5207	4296	gi1161381	IcAB [Staphylococcus epidermidis]	82	61	912
315	3	4864	2855	gi143397	quinol oxidase [Bacillus subtilis]	82	67	2010
321	10	8520	7945	gi1143981	ORF5; This ORF includes a region (aa23-103) containing a potential con- sulphur centre homologous to a region of Rhodospirillum rubrum and Chromatium vinosum; putative [Bacillus stearothermophilus] pir PQ0295 PQ0299 hypothetical protein 5 (gldA 3' region) -	82	62	576
331	3	1055	1342	gi1436574	ribosomal protein L1 [Bacillus subtilis]	82	71	288
370	2	262	616	gi11303793	YqeL [Bacillus subtilis]	82	59	357
404	4	3053	4024	gi11303821	YqfE [Bacillus subtilis]	82	68	972
405	4	4440	3073	gi11303913	YqhX [Bacillus subtilis]	82	67	1368
436	3	4096	2884	gi1149521	cryptophan synthase beta subunit [Lactococcus lactis] pir S35129 S35129 cryptophan synthase (EC 4.2.1.20) beta chain - lactococcus lactis subsp. lactis	82	67	1233

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
441	4	3394	2573	gi 142952	glyceraldehyde-3-phosphate dehydrogenase (Bacillus stearothermophilus)	82	67	822
444	112	10415	11227	gi 1204354	spore germination and vegetative growth protein (Haemophilus influenzae)	82	67	813
446	1	3	191	gi 143387	aspartate transcarbamylase (Bacillus subtilis)	82	66	189
462	3	1007	1210	gi 142521	deoxyribodipyrimidine photolyase (Bacillus subtilis) p1r1A371921A37192 uvrB protein - Bacillus subtilis ap14951[UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.	82	64	204
537	1	1560	784	gi 853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)	82	61	777
680	2	407	700	gi 426472	lacE gene product (Staphylococcus carnosus)	82	69	294
724	2	565	386	gi 143373	phosphoribosyl aminoimidazole carboxy formyl ornyltransferase/inosine monophosphate cyclohydrolase (Pur-H3J) Bacillus subtilis	82	68	180
763	1	422	213	gi 467458	cell division protein (Bacillus subtilis)	82	35	210
818	1	564	283	gi 1064787	function unknown (Bacillus subtilis)	82	69	282
858	1	175	1176	gi 143043	uroporphyrinogen decarboxylase (Bacillus subtilis) p1r1B47045[B47045 uroporphyrinogen decarboxylase (EC 4.1.1.37) - acillus subtilis	82	71	1002
895	1	3	599	gi 1027507	ATP binding protein (Borrelia burgdorferi)	82	72	597
939	1	10	399	gi 143795	transfer RNA-Tyr synthetase (Bacillus subtilis)	82	60	390
961	1	1	306	gi 577647	gamma-hemolysin (Staphylococcus aureus)	82	69	306
1192	1	307	155	gi 146974	NH3-dependent NAD synthetase (Escherichia coli)	82	71	153
1317	1	49	375	gi 407908	Elisac (Staphylococcus xylosus)	82	72	327
1341	1	1	150	gi 39962	ribosomal protein L35 (AA 1-66) (Bacillus stearothermophilus)	82	68	150
1990	2	567	349	gi 534855	ATPase subunit epsilon (Bacillus stearothermophilus) ap1P42009[ATPE_BACST ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).	82	47	219
3024	1	45	224	gi 467402	unknown (Bacillus subtilis)	82	64	180
3045	1	276	139	gi 467335	ribosomal protein L9 (Bacillus subtilis)	82	60	138
3045	2	558	400	gi 467335	ribosomal protein L9 (Bacillus subtilis)	82	82	159
3091	1	474	238	gi 499335	lacA protein (Staphylococcus carnosus)	82	78	237
3107	1	416	210	gi 546918	orfV 3' of comK (Bacillus subtilis, E26, Peptide Partial, 140 aa) p1r1S43612[S43612 hypothetical protein Y - Bacillus subtilis ap1P40398][YHXD_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFV) FRAGMENT).	82	64	207

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4332	1	2	319	gi 42086	nitrate reductase alpha subunit [Escherichia coli] p P09152 NARG_ECOLI RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). [SUB 2-1247]	82	75	318
23	3	3275	2574	gi 1199573	spasB [Spingomonas sp.]	81	64	702
42	1	638	321	gi 466778	lysine specific permease [Escherichia coli]	81	59	318
48	5	4051	4350	gi 1045937	N. genitalium predicted coding region NG216 [Mycoplasma genitalium]	81	62	300
51	4	1578	2579	p t S16649 S166	dciAC protein - Bacillus subtilis	81	55	1002
53	2	354	1494	gi 1303961	YqjJ [Bacillus subtilis]	81	67	1131
53	8	9419	7971	gi 146930	6-phosphogluconate dehydrogenase [Escherichia coli]	81	66	1449
54	9	10757	10119	gi 143016	permease [Bacillus subtilis]	81	65	639
54	10	13360	11786	gi 143015	gluconate kinase [Bacillus subtilis]	81	64	1575
57	17	13983	13366	p t A25805 A258	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi 1222302	Nifu-related protein [Haemophilus influenzae]	81	54	492
86	1	745	374	gi 414077	jpa-93d gene product [Bacillus subtilis]	81	70	372
103	6	6438	4861	gi 971342	nitrate reductase beta subunit [Bacillus subtilis] sp P42176 NARL_BACSU NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4).	81	64	1578
120	15	10845	12338	gi 1524392	GbsA [Bacillus subtilis]	81	67	1494
128	5	3676	4413	gi 143319	triase phosphate isomerase [Bacillus megaterium]	81	64	738
131	9	10308	9280	gi 299163	alanine dehydrogenase [Bacillus subtilis]	81	68	1039
143	6	6088	5471	gi 439619	[Salmonella typhimurium] IS200 insertion sequence from SAR417, art1a1.1. gene product [Salmonella typhimurium]	81	61	618
169	1	43	825	gi 697795	30S ribosomal protein (Padiococcus acidilactici) sp P49668 RS2_PEDAC 30S RIBOSOMAL PROTEIN S2.	81	65	783
230	1	450	226	gi 1125826	short region of weak similarity to tyrosine-protein kinase receptors in a fibronectin type III-like domain [Caenorhabditis elegans]	81	54	225
233	5	2000	2677	gi 467404	unknown [Bacillus subtilis]	81	63	678
241	2	3081	2149	gi 16510	succinate-CoA ligase (GDP-forming) [Arabidopsis thaliana] tr S10579 S10579 succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) pha chain - Arabidopsis thaliana (fragment)	81	69	933
256	1	1	981	nr S09411 S094	spoIIIE protein - Bacillus subtilis	81	65	981
259	3	3752	2691	sp P28367 RP2_B	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RP-2) (FRAGMENT).	81	65	1062

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
275	2	1728	3581	gi1726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	81	68	1854
285	1	1466	735	gi11204844	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	81	63	732
296	1	99	1406	gi1467328	adenylosuccinate synthetase [Bacillus subtilis]	81	67	1308
302	9	5590	5889	gi1147485	cueA [Escherichia coli]	81	64	300
317	2	1137	1376	gi1354961	resolvase [Transposon Tn917]	81	54	240
343	2	1034	1342	gi1405955	lyeD [Escherichia coli]	81	60	309
360	2	1404	2471	gi11204570	aspartyl-tRNA synthetase [Haemophilus influenzae]	81	67	1068
364	5	6251	5706	gi11204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	81	63	546
372	2	1707	1135	gi1467416	unknown [Bacillus subtilis]	81	65	573
392	1	43	603	pirS0941 S094	spoIIIE protein - Bacillus subtilis	81	65	561
404	9	5252	6154	gi1606745	Bex [Bacillus subtilis]	81	65	903
426	2	1727	1119	gi139453	Manganese superoxide dismutase [Bacillus caldotenax] ir[S22053 S22053 superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus licheniformis]	81	66	609
480	7	1653	5889	pi11370H3 C170	hypothetical protein II (ompl 3' region) - Salmonella typhimurium (fragment)	81	57	217
625	3	1105	2070	gi1262360	protein kinase PknB [Mycobacterium leprae]	81	56	966
754	2	504	1064	gi11303902	YqhU [Bacillus subtilis]	81	71	561
842	1	86	430	gi11405446	transketolase [Bacillus subtilis]	81	68	345
953	1	798	400	gi11205429	dipeptide transport ATP-binding protein [Haemophilus influenzae]	81	57	399
961	2	252	401	gi1487686	synergohymenotropic toxin [Staphylococcus intermedius] pir[S44944 S44944 synergohymenotropic toxin - Staphylococcus intermedius]	81	72	150
1035	1	1	189	gi11046138	M. genitalium predicted coding region MG433 [Mycoplasma genitalium]	81	43	189
1280	1	670	449	gi1559164	helicase [Autographa californica nuclear polyhedrosis virus] sp[P24307 V143_NPVAC HELICASE]	81	43	222
3371	1	68	241	gi1132245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	81	62	174
3715	1	475	239	gi1537137	ORF_1388 [Escherichia coli]	81	58	237
3908	1	2	325	gi1439619	[Salmonella typhimurium IS200 insertion sequence from SARA17, attial.]. gene product [Salmonella typhimurium]	81	68	324
3940	1	3	401	in1296464	ATPase [Lactococcus lactis]	81	69	399

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3954	1	1	318	gi1224069	amidase [Moraxella catarrhalis]	81	68	318
4049	1	337	170	gi1603768	HutC protein, imidazolone-5-propanone hydrolase [Bacillus subtilis] gi1603768 HutC protein, imidazolone-5-propanone hydrolase [Bacillus subtilis]	81	68	168
4209	1	1	324	gi1403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir[S37251]S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	81	58	324
4371	1	627	322	gi1216677	indolepyruvate decarboxylase [Enterobacter cloacae] pir[S16013] indolepyruvate decarboxylase (EC 4.1.1.-) - nterobacter cloacae	81	72	306
4387	1	19	228	gi1460689	TWU [Thermactinomyces vulgaris]	81	59	210
4391	1	581	306	gi11524193	unknown [Mycobacterium tuberculosis]	81	67	276
4425	1	3	341	gi1143015	glucuronate kinase [Bacillus subtilis]	81	66	339
9	2	1593	847	gi1064786	function unknown [Bacillus subtilis]	80	62	747
17	1	544	311	gi1559164	helicase [Autographa californica nuclear polyhedrosis virus] sp[P24307]v143 NPVAC HELICASE	80	40	234
45	2	1159	2448	gi11109684	ProV [Bacillus subtilis]	80	63	1290
45	5	4032	4733	gi11109687	ProZ [Bacillus subtilis]	80	55	702
54	8	10266	9502	gi1563952	glucuronate permease [Bacillus licheniformis]	80	62	765
62	12	8852	7545	gi1854655	Na/H antiporter system [Bacillus alcalophilus]	80	62	1308
62	14	8087	8683	gi1559713	ORP [Homo sapiens]	80	68	597
67	16	13781	14122	gi1305002	ORF.F356 [Escherichia coli]	80	65	342
70	13	11495	10296	gi11303995	YqkN [Bacillus subtilis]	80	64	1200
98	9	6336	7130	gi1467428	unknown [Bacillus subtilis]	80	68	795
98	10	7294	7833	gi1467430	unknown [Bacillus subtilis]	80	64	540
98	11	7820	8737	gi1467431	high level kasamycin resistance [Bacillus subtilis]	80	61	918
109	16	14154	14813	gi1580875	ipa-57d gene product [Bacillus subtilis]	80	63	660
112	15	14294	16636	gi1072361	pyruvate-formate-lyase [Clostridium pasteurianum]	80	65	2343
139	1	1448	726	gi1506699	CapC [Staphylococcus aureus]	80	58	723
139	2	2179	1448	gi1506698	CapB [Staphylococcus aureus]	80	59	732
174	4	3271	2870	gi1146242	aspartate 1-decarboxylase [Bacillus subtilis]	80	61	402

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
177	3	2102	2842	gi1467385	unknown [Bacillus subtilis]	80	70	741
184	6	6124	5912	gi1161953	85-KDa surface antigen (Trypanosoma cruzi)	80	46	213
186	4	5368	3875	gi1289282	glutaryl-tRNA synthetase [Bacillus subtilis]	80	65	1494
205	20	15796	15140	gi140103	ribosomal protein L4 [Bacillus stearothermophilus]	80	66	657
207	1	140	1315	gi1460259	enolase [Bacillus subtilis]	80	67	1176
211	3	1078	1590	gi1410131	ORP7 [Bacillus subtilis]	80	61	513
235	2	1962	2255	gi1143787	valyl-tRNA synthetase [Bacillus stearothermophilus] sp P19311 SVV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE (VALRS).	80	55	294
239	1	1	1263	gi1143000	proton glutamate symport protein [Bacillus stearothermophilus] p J56247 J56247 glutamate/aspartate transport protein - Bacillus stearothermophilus	80	59	1263
272	5	2724	2461	gi1709993	hypothetical protein [Bacillus subtilis]	80	54	264
301	3	1446	1111	gi1467418	unknown [Bacillus subtilis]	80	58	336
310	4	5697	4501	gi1177686	acuc gene product [Staphylococcus xylosus]	80	67	1197
310	6	5258	7006	gi1348053	acetyl-CoA synthetase [Bacillus subtilis]	80	67	1749
310	7	7410	9113	gi1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	80	67	1704
325	3	1114	1389	gi1310325	outer capsid protein [Rotavirus sp.]	80	40	276
337	1	1268	636	gi1537049	ORF_0470 [Escherichia coli]	80	55	633
374	2	929	1228	gi1405448	YneF [Bacillus subtilis]	80	70	300
375	5	3062	3331	gi1467448	unknown [Bacillus subtilis]	80	68	270
388	1	267	587	gi1066791	function unknown [Bacillus subtilis]	80	65	321
394	1	9	659	gi1304976	matches PS00017: ATP_GTP_A and PS00301: EFPACTOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins [Escherichia coli]	80	65	651
456	1	625	1263	gi1146183	putative [Bacillus subtilis]	80	65	639
475	1	1	654	gi1288269	beta-fructofuranosidase [Staphylococcus xylosus]	80	66	654
544	2	1449	2240	gi1529754	speC [Streptococcus pyogenes]	80	50	792
622	4	1623	1871	gi11483545	unknown [Mycobacterium tuberculosis]	80	65	249
719	1	1	1257	gi1066791	function unknown [Bacillus subtilis]	80	68	1257
739	1	107	838	gi1666983	putative ATP binding subunit [Bacillus subtilis]	80	61	733

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
745	2	581	414	gi 1511600	coenzyme PQQ synthesis protein III [Methanococcus jannaschii]	80	61	168
822	1	17	679	gi 410141	ONP17 [Bacillus subtilis]	80	68	663
827	2	901	836	gi 1205301	leukotoxin secretion ATP-binding protein [Haemophilus influenzae]	80	54	156
1044	1	3	149	gi 60632	VP2 [Marburg virus]	80	55	147
1220	2	571	413	pir A61072 EPSC	galldarmin precursor - Staphylococcus gallinarum	80	74	159
2519	1	75	275	gi 147556	[dp] [Escherichia coli]	80	45	201
2947	1	503	279	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	80	62	225
3120	1	2	226	gi 517205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	80	65	225
3191	1	284	148	gi 151259	HMG-CoA reductase (EC 1.1.1.86) [Pseudomonas mvalonii] pir A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) [Pseudomonas sp.]	80	59	147
3560	2	285	434	gi 217110	photosystem I core protein B [Synechococcus vulcanus]	80	70	150
3655	1	47	346	gi 415855	deoxyribose aldolase [Mycoplasma hominis]	80	56	300
3658	2	324	584	gi 551531	2-nitropropane dioxygenase [Williopsis saturnus]	80	54	261
3769	1	798	400	gi 133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	399
3781	1	692	348	gi 166412	NADH-glutamate synthase [Medicago sativa]	80	62	345
3988	1	48	287	gi 1204696	fructose-6-phosphate IIBC component [Haemophilus influenzae]	80	69	240
4030	1	571	287	gi 1009166	respiratory nitrate reductase [Bacillus subtilis]	80	60	285
4092	1	547	275	gi 1370207	orf6 [Lactobacillus sake]	80	69	273
4103	1	680	342	gi 39956	ilc1c [Bacillus subtilis]	80	65	339
4231	1	692	348	gi 289287	UDP-glucose pyrophosphorylase [Bacillus subtilis]	80	65	345
4265	1	595	299	gi 603768	HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis]	80	63	297
4504	1	498	250	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	249
2	6	5998	6798	gi 535351	CODY [Bacillus subtilis]	79	63	801
4	7	8295	7057	gi 603768	HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis]	79	64	1245
25	6	5273	5515	pir A36728 A367	acyl carrier protein - Rhizobium meliloti	79	65	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
59	2	1173	1424	gi 147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	79	75	252
60	1	1	204	gi 666115	orf1 upstream of glucose kinase [Staphylococcus xylosus] pir S52351 S52351 hypothetical protein 1 - Staphylococcus xylosus	79	60	204
81	1	3002	1590	gi 466882	[ppa1; 81896_C2_189 Mycobacterium leprae]	79	64	1413
85	7	7023	6505	gi 143364	[phosphoribosyl aminoimidazole carboxylase 1 (PUR-E) [Bacillus subtilis]	79	60	519
89	6	5660	4554	gi 144906	product homologous to E. coli thioesterase reductase: J.Biol.Chem. 1988; 263:9015-9019, and to F32a protein of alkyl hydroperoxide oxidase from S. typhimurium: J. Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	79	35	1107
102	11	7489	8571	gi 143093	ketol-acid reductoisomerase [Bacillus subtilis] sp p37253 ILVC_BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOHYDROXY-ACID ISOMEROREDUCTASE) (ALPHA-KETO-BETA-HYDROXYLACIL EDUCTOISOMERASE)	79	64	1083
102	14	11190	12563	gi 149428	[putative [Lactococcus lactis]	79	65	1374
127	9	7792	9372	gi 458688	[PrfC/Rf3] [Dichalobacter nodosus]	79	68	1581
139	3	2540	1983	gi 506697	[CapA [Staphylococcus aureus]	79	55	558
144	2	1644	1156	gi 1498296	[peptide methionine sulfoxide reductase (Streptococcus pneumoniae)]	79	47	489
144	2	529	1098	gi 467457	[hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi 467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]	79	59	570
150	1	965	591	gi 755602	[unknown [Bacillus subtilis]	79	61	375
176	1	1039	587	gi 297874	[fructose-bisphosphate aldolase [Staphylococcus carnosus] pir A69943 A69943 fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TN300)]	79	65	453
186	7	7584	6874	gi 1314298	[ORF5; putative Sms protein; similar to Sms proteins from Haemophilus influenzae and Escherichia coli [Listeria monocytogenes]	79	64	711
205	16	8887	8498	gi 1044980	[ribosomal protein L18 [Bacillus subtilis]	79	70	390
211	1	1	519	gi 1303994	[YqkH [Bacillus subtilis]	79	62	519
223	2	4183	2801	gi 488430	[alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1383
243	8	8915	7896	gi 580883	[ipa-88d gene product [Bacillus subtilis]	79	60	1020
279	4	3721	4329	gi 413930	[ipa-6d gene product [Bacillus subtilis]	79	59	609
300	1	11	1393	gi 403372	[glycerol 3-phosphate permease [Bacillus subtilis]	79	62	1383
307	3	2930	1935	gi 950062	[hypothetical yeast protein 1 [Mycoplasma capricolum] pir S48578 S48578 hypothetical protein - Mycoplasma capricolum SCC3] (fragment)	79	60	996

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
352	6	10106	8886	gi 216854	[Pseudomonas chlororaphis]	79	59	1221
412	1	1153	578	gi 143177	[putative Bacillus subtilis]	79	51	576
481	3	621	1124	gi 786163	[Ribosomal Protein L10 [Bacillus subtilis]]	79	66	504
516	1	702	352	gi 805090	[NarP [Lactococcus lactis]]	79	48	351
525	2	2457	1426	gi 143371	[phosphoribosyl aminimidazole synthetase (pur-H) [Bacillus subtilis]]	79	61	1032
				pir H29326 AJ850	[phosphoribosylformylglycinamide cyclase EC 6.3.3.1] - Bacillus subtilis			
534	4	3448	2825	gi 1370207	[orf6 [Lactobacillus sake]]	79	67	624
570	1	2	421	gi 476160	[arginine permease substrate-binding subunit [Listeria monocytogenes]]	79	61	420
645	8	2663	3241	gi 153898	[transport protein [Salmonella typhimurium]]	79	62	579
683	1	75	374	gi 1064795	[function unknown [Bacillus subtilis]]	79	62	714
816	3	4700	3987	gi 1407784	[orf-1; novel antigen [Staphylococcus aureus]]	79	61	399
2929	1	3	401	gi 1524397	[glycine betaine transporter Opud [Bacillus subtilis]]	79	58	156
2937	1	357	202	pir S52915 S529	[nitrate reductase alpha chain - Bacillus subtilis (fragment)]	79	72	384
2940	1	768	385	gi 149429	[putative [Lactococcus lactis]]	79	61	285
2946	1	570	286	gi 143267	[2-oxoglutarate dehydrogenase (odhA; EC 1.2.4.2) [Bacillus subtilis]]	79	59	210
2999	1	3	212	gi 710020	[nitrite reductase (nirB) [Bacillus subtilis]]	79	61	183
3022	1	514	332	gi 430686	[3-phosphoglycerate kinase [Thermotoga maritima]]	79	60	312
3064	1	3	314	gi 1204436	[pyruvate formate-lyase [Haemophilus influenzae]]	79	56	219
3083	1	2	220	gi 1149662	[hlyD gene product [Clostridium perfringens]]	79	55	291
3126	1	701	411	gi 1339950	[large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]]	79	59	282
3181	1	607	326	gi 1339950	[large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]]	79	63	474
3345	1	3	476	gi 871784	[Clp-like ATP-dependent protease binding subunit [Bos taurus]]	79	71	267
3718	1	536	270	pir C36889 C368	[leuB protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)]	79	64	243
3724	2	159	401	gi 1009366	[Respiratory nitrate reductase [Bacillus subtilis]]	79	65	237
3836	1	608	312	gi 1524193	[unknown [Mycobacterium tuberculosis]]	79	56	333
3941	1	2	334	gi 415855	[deoxyribose aldolase [Mycoplasma hominis]]	79	63	339
4113	1	3	341	gi 143015	[gluconate kinase [Bacillus subtilis]]	79	63	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4501	1	406	209	gi11022726	unknown [Staphylococcus haemolyticus]	79	66	198
4612	1	2	238	gi1460689	TVG [Thermactinomyces vulgaris]	79	58	237
2	1	2	1213	gi1520753	DNA topoisomerase I [Bacillus subtilis]	78	64	1212
8	2	2266	1220	gi1216151	DNA polymerase (gene L; ttc start codon) [Bacteriophage SPO2] gi1579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] p1r[A21498]DJBP52 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	78	72	1047
9	2	1340	1089	gi11064787	function unknown [Bacillus subtilis]	78	57	252
32	8	6803	7702	gi1146974	NHJ-dependent NAD synthetase [Escherichia coli]	78	63	900
36	4	2941	3138	gi1290503	glutamate permease [Escherichia coli]	78	53	198
53	135	117684	116221	gi11303941	YqjV [Bacillus subtilis]	78	58	1464
57	16	10520	12067	gi11072418	glcA gene product [Staphylococcus carnosus]	78	65	1548
66	7	6798	5812	gi11212729	YqjH [Bacillus subtilis]	78	67	987
67	4	4029	4376	gi1466612	nika [Escherichia coli]	78	71	348
91	9	110058	10942	gi1467380	stage 0 sporulation [Bacillus subtilis]	78	50	885
102	12	8574	10130	gi1149426	putative [Lactococcus lactis]	78	61	1557
112	6	3540	4463	gi1854234	cymC gene product [Klebsiella oxytoca]	78	56	924
124	2	1888	1061	gi1405622	unknown [Bacillus subtilis]	78	60	828
130	3	1805	2260	gi11256636	putative [Bacillus subtilis]	78	71	456
133	1	751	377	gi1168060	lamb [Emicella nidulans]	78	59	375
166	4	7125	6163	gi1451216	Mannosephosphate isomerase [Streptococcus mutans]	78	63	963
186	1	1586	795	gi1289284	cysteinyI-cRNA synthetase [Bacillus subtilis]	78	63	792
195	4	2749	2315	gi11353874	unknown [Rhodobacter capsulatus]	78	58	435
199	3	4279	3623	gi1143525	succinate dehydrogenase cytochrome b-558 subunit [Bacillus subtilis] p1r[A29843]DESSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis	78	57	657
199	4	7209	5557	gi1142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] p1r[A27192]A27192 uvrB protein - Bacillus subtilis sp116931[uvrC_BACSU EXCINUCLEASE ABC SUBUNIT C.	78	62	1653
223	3	3831	3523	gi1139596	[Escherichia coli t5200 insertion sequence from ECOR63, partial.1, ene product [Escherichia coli]	78	47	309

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
299	4	1865	2149	gi 467439	temperature sensitive cell division [Bacillus subtilis]	78	62	285
321	9	7734	7315	gi 142979	ORF3 is homologous to an ORF downstream of the spot gene of E.coli; RF3 [Bacillus stearothermophilus]	78	55	420
352	4	3714	3944	gi 149050	actin 1 [Pneumocystis carinii]	78	42	231
352	5	7592	6093	gi 903587	NADH dehydrogenase subunit 5 [Bacillus subtilis] sp P39755 NDH2_BACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 51	78	58	1500
376	1	2	583	gi 551693	dethiobiotin synthase [Bacillus sphaericus]	78	34	582
424	2	1595	1768	gi 1524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	78	68	174
450	1	1914	988	gi 1030068	NAD(P)H oxidoreductase, isoflavone reductase homologue [Solanum tuberosum]	78	63	927
558	1	762	562	gi 1511588	bifunctional protein [Methanococcus jannaschii]	78	60	201
670	3	1152	1589	gi 1122759	unknown [Bacillus subtilis]	78	64	438
714	1	64	732	gi 143460	37 kd minor sigma factor (rpoF, sigB; ttg start codon) [Bacillus subtilis]	78	57	669
814	1	3	368	gi 1377833	unknown [Bacillus subtilis]	78	59	366
981	1	1381	692	gi 143802	CerC2 [Bacillus subtilis]	78	64	690
995	2	978	727	gi 296947	uridine kinase [Escherichia coli]	78	64	252
1045	1	3	401	gi 1407784	orf-1; novel antigen [Staphylococcus aureus]	78	61	399
1163	2	368	186	gi 410117	ulaminopimelate decarboxylase [Bacillus subtilis]	78	54	183
2191	1	794	399	gi 215098	excisionase [Bacteriophage 156e]	78	65	396
2933	1	2	181	gi 120436	pyruvate formate-lyase [Haemophilus influenzae]	78	73	180
3041	2	129	317	gi 624632	ClcL [Escherichia coli]	78	53	189
3581	1	105	401	gi 763186	3-ketoacyl-CoA thiolase [Saccharomyces cerevisiae]	78	55	297
3709	1	3	230	gi 460689	rvc [Thermactinomyces vulgaris]	78	58	228
3974	1	528	265	gi 538839	unknown [Bacillus subtilis]	78	65	264
3980	1	3	401	gi 39956	ITIC [Bacillus subtilis]	78	62	399
4056	1	647	334	gi 1256635	[dehydroxy-acid dehydratase [Bacillus subtilis]	78	55	294
4114	1	630	316	gi 509172 509	hypothetical protein - Trypanosoma brucei	78	62	315
4185	1	3	379	gi 133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	78	58	177

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4235	1	655	329	gi 558839	unknown [Bacillus subtilis]	78	60	327
4352	1	541	302	gi 603768	HutI protein, imidazole-5-propionate hydrolase [Bacillus subtilis]	78	63	240
4368	1	612	307	gi 1351678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	78	59	306
4461	1	428	216	gi 1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	78	36	213
4530	1	474	238	gi 39956	IGIC [Bacillus subtilis]	78	65	237
3	2	2969	2073	gi 1109684	ProV [Bacillus subtilis]	77	56	897
12	2	2426	1965	gi 467335	ribosomal protein L9 [Bacillus subtilis]	77	59	462
27	1	2	388	gi 1212728	YghI [Bacillus subtilis]	77	63	387
39	2	590	1252	gi 40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-604) [Bacillus subtilis]	77	60	663
42	6	2704	2931	gi 606241	30S ribosomal subunit protein S14 [Escherichia coli] sp P02370 SS14_ECOLI	77	65	228
46	18	15459	16622	gi 297798	mitochondrial formate dehydrogenase precursor [Solanum tuberosum]	77	55	1164
100	4	4562	4002	gi 1340128	ORF1 [Staphylococcus aureus]	77	54	561
102	8	5378	5713	gi 1311482	acetolactate synthase [Thermus aquaticus]	77	57	336
109	7	4742	5383	gi 710637	unknown [Bacillus subtilis]	77	56	642
117	1	2	1228	gi 1237015	ORF4 [Bacillus subtilis]	77	53	1227
124	10	8323	7688	gi 405819	thymidine kinase [Bacillus subtilis]	77	63	636
147	3	1146	985	gi 849027	hypothetical 15.9-kDa protein [Bacillus subtilis]	77	37	162
152	10	7354	7953	gi 1205583	spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]	77	55	600
169	2	1004	1282	gi 473825	'elongation factor EF-Ts' [Escherichia coli]	77	58	279
184	2	380	1147	gi 216314	esterase [Bacillus stearothermophilus]	77	60	768
189	7	3296	3868	gi 851809	ORF3 [Clostridium perfringens]	77	48	573
193	1	132	290	gi 1303788	YqoH [Bacillus subtilis]	77	54	159
195	8	8740	8414	gi 1499620	M. jannaschii predicted coding region M0798 [Methanococcus jannaschii]	77	44	327
205	8	5428	5204	gi 216340	ORF for adenylate kinase [Bacillus subtilis]	77	61	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
205	29	14795	14502	gi 786155	Ribosomal Protein L23 (Bacillus subtilis)	77	62	294
211	5	1908	2084	gi 410132	OMP8 (Bacillus subtilis)	77	47	177
217	5	3478	4416	gi 496254	[fibronectin/fibrinogen-binding protein (Streptococcus pyogenes)]	77	54	939
222	1	267	998	gi 407784	[orf-1; novel antigen (Staphylococcus aureus)]	77	57	732
233	2	1819	1346	gi 467408	[unknown (Bacillus subtilis)]	77	61	474
243	3	2661	2299	gi 516155	[unconventional myosin (Sus scrofa)]	77	32	363
299	1	68	769	gi 467436	[unknown (Bacillus subtilis)]	77	54	702
301	4	1468	1283	gi 950071	[ATP-bind, pyruvate kinase (Mycoplasma capricolum) pir S48605[S48605 hypothetical protein - Mycoplasma capricolum SOC3] (fragment)]	77	48	186
302	5	2741	3211	gi 508980	[pheB (Bacillus subtilis)]	77	57	471
302	7	3835	4863	gi 147783	[ruvB protein (Escherichia coli)]	77	60	1039
307	9	5402	4797	gi 1070015	[protein-dependent (Bacillus subtilis)]	77	60	606
312	1	99	1391	gi 143165	[malic enzyme (EC 1.1.1.38) (Bacillus stearothermophilus) pir A33307 DBSXS maleate dehydrogenase oxaloacetate-decarboxylating (EC 1.1.1.38) - Bacillus stearothermophilus]	77	62	1293
312	2	1361	2443	gi 1399855	[carboxyltransferase beta subunit (Symbiobacterium PCC7942)]	77	58	903
321	5	5666	4596	gi 39844	[fumarate (citG) (aa 1-462) (Bacillus subtilis)]	77	65	1071
354	1	47	568	gi 1154634	[ymaB (Bacillus subtilis)]	77	57	522
365	1	2	1021	gi 143374	[phosphoribosyl glycineamide synthetase (PUR-D; gtc start codon) Bacillus subtilis]	77	62	1020
374	1	1	708	gi 1405446	[transketolase (Bacillus subtilis)]	77	61	708
385	1	1128	565	gi 533099	[endonuclease III (Bacillus subtilis)]	77	63	564
392	2	594	1940	gi 556014	[UDP-N-acetyl muranate-alanine ligase (Bacillus subtilis) sp P40778 KURC_BACSU_UDP-N-ACETYLURANATE--ALANINE LIGASE (EC 3.2.8) (UDP-N- ACETYLURANATE--L-ALANINE SYNTHETASE) (FRAGMENT)]	77	65	1347
405	5	4079	3570	gi 1303912	[yqhW (Bacillus subtilis)]	77	64	510
487	4	1302	1472	gi 432427	[ORF1 gene product (Acinetobacter calcoaceticus)]	77	48	171
522	1	2	562	pir A01179 Syns	[tyrosine--trna ligase (EC 6.1.1.1) - Bacillus stearothermophilus]	77	63	561

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	match gene name	% sim	% ident	length (nt)
523	2	1587	1351	gi 1387979	44% identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D64006.CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	77	48	237
536	2	983	612	gi 143366	adenylosuccinate lyase (PUB-9) [Bacillus subtilis] pir C29326 W2B5DS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	77	61	372
548	2	339	872	gi 143387	aspartate transcarbamylase [Bacillus subtilis]	77	56	534
597	1	2	481	gi 1904198	hypothetical protein [Bacillus subtilis]	77	33	480
633	2	1747	1313	gi 187577	ORF1A [Bacillus subtilis]	77	64	435
642	1	85	360	gi 46971	epiP gene product [Staphylococcus epidermidis]	77	61	276
659	1	125	1219	gi 1072381	glutamyl-aminopeptidase [Lactococcus lactis]	77	62	1095
670	4	1587	1820	gi 1122760	unknown [Bacillus subtilis]	77	58	234
789	1	2	391	gi 1377823	aminopeptidase [Bacillus subtilis]	77	65	390
815	1	10	573	gi 1303861	VqgN [Bacillus subtilis]	77	49	564
899	1	1	225	gi 1204844	H. influenzae predicted coding region H0594 [Haemophilus influenzae]	77	55	225
1083	1	3	188	gi 460828	B969 [Saccharomyces cerevisiae]	77	66	186
1942	1	415	209	gi 160047	P101/acidic basic repeat antigen [Plasmodium falciparum] pir A29232 A29232 101K malaria antigen precursor - Plasmodium falciparum (strain Camp)	77	38	207
2559	1	1	171	gi 1499034	M. Jannaschii predicted coding region MJ0255 [Methanococcus jannaschii]	77	61	171
2933	2	243	401	gi 42370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir S01788 S01788 formate C-acetyltransferase (EC 2.3.1.54) - Escherichia coli	77	72	159
2966	1	56	292	gi 1524397	glycine betaine transporter OpuD [Bacillus subtilis]	77	45	237
2976	1	614	309	gi 40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p P23129 O001_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)	77	60	306
2979	2	678	400	gi 1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	77	61	279
2988	1	601	377	gi 438465	Probable operon with orfP. Possible alternative initiation codon, asa 2151-2153. Homology with acetyltransferases; putative Bacillus subtilis	77	55	225
2990	1	331	167	gi 142562	ATP synthase epsilon subunit [Bacillus megaterium] pir B28599 PWBSEH H+transporting ATP synthase (EC 3.6.1.34) psilon chain - Bacillus megaterium	77	63	165
3032	1	3	389	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	77	56	387
3057	1	1	195	gi 468764	mocR gene product [Rhizobium meliloti]	77	50	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4008	1	726	400	gi 603768	Hut1 protein, imidazole-5-propionate hydrolase [Bacillus subtilis] gi 603768 Hut1 protein, imidazole-5-propionate hydrolase Bacillus subtilis	77	52	327
4048	1	703	386	gi 216278	Gramicidin S synthetase 1 [Bacillus brevis]	77	55	318
4110	1	3	368	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	366
4115	1	1	348	gi 517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	77	65	348
4225	1	590	297	gi 1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	77	60	294
4611	2	494	327	gi 508979	GTP-binding protein [Bacillus subtilis]	77	57	168
4668	1	361	182	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	180
25	1	2	1627	gi 1150620	HsaA [Streptococcus pneumoniae]	76	58	1626
38	5	1488	2537	pir A43577 A435	regulatory protein pfor - Clostridium perfringens	76	57	1050
52	5	2962	4041	gi 1161061	dioxigenase [Methylobacterium extorquens]	76	62	1080
56	20	27389	27955	gi 467402	unknown [Bacillus subtilis]	76	56	567
57	15	12046	12219	gi 1206040	weak similarity to keratin [Caenorhabditis elegans]	76	40	174
91	2	1062	2261	gi 475715	acetyl coenzyme A acetyltransferase [Clostridium acetobutylicum]	76	57	1200
98	2	818	1624	gi 467422	unknown [Bacillus subtilis]	76	62	807
98	5	2965	3228	gi 897793	y98 gene product [Pediococcus acidilactici]	76	52	264
98	8	5922	6326	gi 467427	methionyl-tRNA synthetase [Bacillus subtilis]	76	53	405
104	3	1322	1685	gi 216151	DNA polymerase (gene L; ttg start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pir A21498 DJ8PS2 DNA-directed DNA polymerase [EC 2.7.7.1] - phage P02	76	63	564
124	9	8134	7055	gi 853776	peptide chain release factor 1 [Bacillus subtilis] pir S55437 S55437 peptide chain release factor 1 - Bacillus ubilis	76	58	1080
164	5	2832	3311	gi 1204976	prolyl-CRNA synthetase [Haemophilus influenzae]	76	53	480
168	2	2617	1841	gi 1177253	putative ATP-binding protein of ABC-type [Bacillus subtilis]	76	58	777
189	2	163	888	gi 467384	unknown [Bacillus subtilis]	76	63	726
235	3	2253	3518	gi 142936	folyl-polyglutamate synthetase [Bacillus subtilis] pir B40646 B40646 folC - Bacillus subtilis	76	53	1266
236	1	335	925	gi 1146197	putative [Bacillus subtilis]	76	54	591
237	8	5323	5541	gi 1279261	F13C3.6 [Caenorhabditis elegans]	76	47	219

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
263	5	5490	4585	gi1510348	[dihydrodipicolinate synthase (Methanococcus jannaschii)]	76	49	906
304	3	1051	1794	gi1666982	[putative membrane spanning subunit (Bacillus subtilis) pir[S52382/S52382] probable membrane spanning protein - Bacillus subtilis]	76	60	744
332	4	3611	4624	gi1143312	[6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) (Bacillus carotermophilus)]	76	56	1014
343	1	2	1036	gi1405956	[yeaE (Escherichia coli)]	76	59	1035
347	1	409	1701	gi1396304	[acetylornithine deacetylase (Escherichia coli)]	76	72	1293
358	1	672	1907	gi1146215	[39.0% identity to the Escherichia coli S1 ribosomal protein; putative (Bacillus subtilis)]	76	58	1236
371	1	1	222	gi1537084	[alternate gene name mgt; CG Site No. 497 (Escherichia coli) pir[S56468/S56468] mgtA protein - Escherichia coli]	76	61	222
379	4	4331	4858	gi1143268	[dihydrolipoyl transsuccinylase (odhA; EC 2.3.1.61) (Bacillus subtilis)]	76	61	528
404	5	4022	4492	gi1303823	[YqfC (Bacillus subtilis)]	76	60	471
411	1	2	307	gi1186025	[ORF YML027w (Saccharomyces cerevisiae)]	76	55	306
472	3	4356	2854	gi11405464	[AlaT (Bacillus subtilis)]	76	57	1503
546	1	273	995	gi1153821	[streptococcal pyrogenic exotoxin type C (speC) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	gi11002520	[MutS (Bacillus subtilis)]	76	61	498
591	1	16	735	gi1885914	[ClpB (Synechococcus sp.)]	76	44	720
602	2	175	798	gi11486422	[OppD homologue (Rhizobium sp.)]	76	52	624
619	2	547	280	gi1330613	[major capsid protein (Human cytomegalovirus)]	76	47	258
660	4	2568	3302	gi1904199	[hypothetical protein (Bacillus subtilis)]	76	55	735
677	1	452	228	gi140177	[spoOF gene product (Bacillus subtilis)]	76	58	225
962	1	24	206	gi1142443	[adenylosuccinate synthetase (Bacillus subtilis) sp[P39726/PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP--ASPARTATE LIGASE)]	76	67	183
978	1	1158	580	gi11511333	[M. jannaschii predicted coding region M3122 (Methanococcus jannaschii)]	76	56	579
997	1	486	244	gi1467154	[No definition line found (Mycobacterium leprae)]	76	38	243
1563	1	529	266	gi1103984	[YqkG (Bacillus subtilis)]	76	52	264
2184	1	361	182	gi1506706	[CapJ (Staphylococcus aureus)]	76	38	180
2572	1	1	387	gi1153898	[transport protein (Salmonella typhimurium)]	76	65	387

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2942	1	29	400	gi 710020	nitrite reductase (nirB) (Bacillus subtilis)	76	59	372
2957	1	377	216	gi 1511251	hypothetical protein (SP-P42404) (Methanococcus jannaschii)	76	47	162
2980	1	554	279	gi 1403464	AlrA (Bacillus subtilis)	76	53	276
3015	1	649	326	gi 408115	ornithine acetyltransferase (Bacillus subtilis)	76	61	324
3124	1	13	174	gi 882705	ORF_0401 (Escherichia coli)	76	65	162
3179	1	3	161	gi 168477	ferredoxin-dependent glutamate synthase (Zea mays) p1r[A38596/A38596] glutamate synthase (ferredoxin) (EC 1.4.7.1) - a1ze	76	53	159
3789	1	2	379	gi 39956	YIGlc (Bacillus subtilis)	76	55	378
3892	1	3	314	gi 1510398	ferritinocellin binding protein (Methanococcus jannaschii)	76	52	312
3928	1	798	400	gi 143016	permease (Bacillus subtilis)	76	59	399
4159	1	757	386	sp P80544 HRSP	METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS)	76	66	372
4204	1	17	331	gi 296464	ATPase (Lactococcus lactis)	76	56	315
4398	1	494	249	gi 987255	Menkes disease gene (Homo sapiens)	76	48	246
4506	1	2	313	gi 216746	D-lactate dehydrogenase (Lactobacillus plantarum)	76	47	312
4546	1	477	247	gi 1339950	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	76	61	231
4596	1	379	191	gi 560027	cellulose synthase (Acetobacter xylinum)	76	70	189
4	5	5257	4337	gi 882532	ORF_0294 (Escherichia coli)	75	59	921
6	1	184	952	gi 40960	ORFase (Escherichia coli)	75	56	789
12	3	5935	3944	gi 467336	unknown (Bacillus subtilis)	75	57	1992
23	18	18272	17310	gi 1296433	O-acetylserine sulphydrylase B (Alcaligenes eutrophus)	75	55	963
25	3	2356	3193	gi 1502419	Plax (Bacillus subtilis)	75	56	1038
36	8	5765	6037	gi 1256517	unknown (Schizosaccharomyces pombe)	75	45	273
46	13	11186	12058	gi 48972	nitrate transporter (Synecococcus sp.)	75	46	873
51	7	3474	3677	gi 143607	sporulation protein (Bacillus subtilis)	75	61	204
53	16	16850	16590	gi 143402	recombination protein (ttg start codon) (Bacillus subtilis) gi 1303923 recN (Bacillus subtilis)	75	51	261
74	3	3572	2568	gi 1204847	ornithine carbamoyltransferase (Haemophilus influenzae)	75	61	1005

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
85	3	4628	3930	gi 143368	phosphoribosylformyl glycine synthetase I (PUR-L, gtc start odon) (Bacillus subtilis)	75	63	699
85	5	5588	4878	gi 143367	phosphoribosyl aminoazole succinocarboxamide synthetase (PUR-C, tgc start codon) (Bacillus subtilis)	75	55	713
85	8	6625	7530	gi 1303916	YqIA (Bacillus subtilis)	75	53	906
87	3	2340	3590	gi 1064813	homologous to ep:PHOR_BACSU (Bacillus subtilis)	75	56	1251
87	6	6084	6896	gi 1064810	function unknown (Bacillus subtilis)	75	61	813
108	2	1844	1503	gi 1001824	hypothetical protein (Synecocystis sp.)	75	51	342
110	3	1748	3727	gi 1147593	putative ppp synthetase (Streptomyces coelicolor)	75	55	1980
110	7	4153	5252	gi 1177251	clwD gene product (Bacillus subtilis)	75	75	900
120	14	11266	10649	gi 1524394	ORF-2 upstream of gbaB operon (Bacillus subtilis)	75	55	618
121	5	2050	4221	gi 1154632	NrdE (Bacillus subtilis)	75	54	2172
124	1	283	143	gi 405622	unknown (Bacillus subtilis)	75	56	141
128	1	81	1139	gi 143316	[gap] gene products (Bacillus megaterium)	75	48	1059
130	8	5760	5903	gi 1256654	54.8% identity with Neisseria gonorrhoeae regulatory protein P110; putative (Bacillus subtilis)	75	62	144
136	2	4480	3185	gi 467403	serY-tRNA synthetase (Bacillus subtilis)	75	54	1296
161	110	5439	5798	gi 1001195	hypothetical protein (Synecocystis sp.)	75	55	360
172	4	3819	2995	gi 755153	ATP-binding protein (Bacillus subtilis)	75	52	825
179	1	2024	1107	gi 143037	porphobilinogen deaminase (Bacillus subtilis)	75	58	918
195	110	9529	9174	gi 1025745 YCPN_	HYPOTHETICAL PROTEIN IN PURB 5'-REGION (ORF-15) (FRAGMENT)	75	60	156
200	4	2605	4596	gi 142440	ATP-dependent nuclease (Bacillus subtilis)	75	56	1992
206	3	6900	5620	gi 1256135	YbbP (Bacillus subtilis)	75	53	1281
216	2	159	389	gi 1052800	unknown (Schizosaccharomyces pombe)	75	58	231
229	1	29	847	gi 1205958	branched chain aa transport system II carrier protein (Haemophilus influenzae)	75	49	819
230	2	518	1714	gi 971337	nitrite extrusion protein (Bacillus subtilis)	75	53	1197
231	1	2340	1123	gi 1002521	MutL (Bacillus subtilis)	75	54	1119
233	3	1314	1859	gi 467405	unknown (Bacillus subtilis)	75	59	546

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
229	1	325	164	gi11511246	[methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]]	75	50	162
232	1	1389	772	gi11511604	[M. jannaschii predicted coding region MJ1651 [Methanococcus jannaschii]]	75	46	618
304	4	1773	2261	gi1205328	[surfacein [Haemophilus influenzae]]	75	55	489
312	3	2437	3387	gi1285621	[undefined open reading frame [Bacillus stearothermophilus]]	75	62	951
312	5	4622	6403	gi11041097	[Pyruvate Kinase [Bacillus psychrophilus]]	75	57	1782
319	1	353	877	gi11212728	[YqjH [Bacillus subtilis]]	75	54	525
320	5	4321	5031	gi11070361	[OMP decarboxylase [Lactococcus lactis]]	75	56	711
320	6	5010	5642	gi1143394	[OMP-PRPP transferase [Bacillus subtilis]]	75	60	633
337	4	1519	2088	gi1487433	[citrate synthase II [Bacillus subtilis]]	75	58	570
394	2	669	1271	gi1304976	[matches P500017: ATP-GTP-A and P500301: EFPACTOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]	75	51	603
423	1	127	570	gi11183839	[unknown [Pseudomonas aeruginosa]]	75	59	444
433	2	1603	1929	gi1149211	[acetolactate synthase [Klebsiella pneumoniae]]	75	63	327
446	2	176	1540	gi1312441	[dihydroorotase [Bacillus caldolyticus]]	75	62	1365
486	1	494	249	gi1149582	[potF gene product [Clostridium perfringens]]	75	55	246
486	1	3	194	gi1143582	[spoIIIEA protein [Bacillus subtilis]]	75	59	792
498	2	824	1504	gi1143328	[phoP protein (put.); putative [Bacillus subtilis]]	75	47	681
499	2	1061	1624	gi1138799	[44% identity over 302 residues with hypothetical protein from Synechocystis sp, accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]]	75	51	564
568	1	641	453	pir13c411013c41	[triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides (SGC3)]	75	50	189
613	2	430	233	gi1330993	[tegument protein [Salvairine herpesvirus 2]]	75	75	198
621	1	1	525	gi1529754	[speC [Streptococcus pyogenes]]	75	43	525
642	5	1609	2474	gi1176401	[EnpG [Staphylococcus epidermidis]]	75	51	666
646	2	454	657	gi1172442	[ribonuclease P [Saccharomyces cerevisiae]]	75	37	204
657	1	3	347	gi1882541	[OMP_0236 [Escherichia coli]]	75	47	345
750	1	1662	832	gi146971	[lepP gene product [Staphylococcus epidermidis]]	75	57	831

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
754	1	2	481	gi 1303901	YohT [Bacillus subtilis]	75	57	480
763	2	563	393	gi 3205145	multidrug resistance protein [Haemophilus influenzae]	75	51	171
775	1	961	482	pir B36889 B368	leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	75	63	480
793	1	1	180	gi 143316	[gap] gene products [Bacillus megaterium]	75	57	180
800	1	318	160	gi 509411	NFRA protein [Azorhizobium caulinodans]	75	34	159
811	1	1117	560	gi 143434	Rho factor [Bacillus subtilis]	75	60	558
940	1	493	329	gi 1276985	arginase [Bacillus caldovelox]	75	50	165
971	2	37	252	gi 1001373	hypothetical protein [Synechocystis sp.]	75	58	216
1059	1	384	232	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	75	67	153
1109	2	219	374	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir A27650 A27650 regulatory protein phor - Bacillus subtilis spi P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)	75	53	156
1268	1	271	137	gi 304135	ornithine acetyltransferase [Bacillus stearothermophilus] spi Q07908 ARG3_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE) (ORNITHINE TRANSACETYLASE) (DATASE) / MINO-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE YNTUA	75	63	135
1500	1	324	163	gi 1205488	exonuclease AQC subunit B [Haemophilus influenzae]	75	57	162
1529	1	798	400	gi 1002521	NutL [Bacillus subtilis]	75	54	399
3010	1	770	387	gi 1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	75	54	384
3105	1	1	180	gi 1041097	Pyruvate Kinase [Bacillus psychrophilus]	75	57	180
3117	1	45	212	gi 899317	peptide synthetase module [Microcystis aeruginosa] pir S4911 S4911 probable amino acid activating domain - [Microcystis aeruginosa (fragment) (SUB 144-528)	75	42	168
3139	2	139	345	gi 145294	adenine phosphoribosyl-transferase [Escherichia coli]	75	66	207
3880	1	618	310	gi 1009166	Respiratory nitrate reductase [Bacillus subtilis]	75	58	309
3911	1	48	401	gi 433991	ATP synthase subunit beta [Bacillus subtilis]	75	68	354
3957	1	2	379	pir B36889 B368	3-isopropylmalate dehydratase (EC 4.2.1.33) chain luuC - Lactococcus lactis subsp. lactis (strain IL1403)	75	65	378
4005	1	5	259	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	75	48	255
4080	1	73	333	gi 415855	deoxyribose aldolase [Mycoplasma hominis]	75	59	261

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4111	1	1	339	gi 149435	putative [Lactococcus lactis]	75	57	339
4136	1	602	303	gi 450688	hscM gene of Ecoprr1 gene product [Escherichia coli] pir[S18437]S18437 hsdM protein - Escherichia coli pir[S09829]S09829 hypothetical protein A - Escherichia coli (SUB 40-520)	75	56	300
4144	1	668	336	gi 48972	nitrate transporter [Synechococcus sp.]	75	49	333
4237	1	664	374	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	75	55	291
4306	2	73	318	gi 294260	major surface glycoprotein [Pneumocystis carinii]	75	68	246
4343	1	715	359	gi 1204652	methylated-DNA-protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	357
4552	1	620	312	gi 298464	ATPase [Lactococcus lactis]	75	55	309
38	9	5776	6126	gi 443793	RupC [Escherichia coli]	74	50	351
50	8	6910	6221	gi 1219988	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	10770	12221	gi 1000451	TrpP [Bacillus subtilis]	74	57	1452
64	2	2266	1622	gi 41015	aspartate-tRNA ligase [Escherichia coli]	74	57	645
66	6	5063	4848	gi 1212729	YqhJ [Bacillus subtilis]	74	47	216
67	18	14334	14897	gi 1510631	endoglucanase [Methanococcus jannaschii]	74	52	564
102	15	12561	13136	gi 149429	putative [Lactococcus lactis]	74	67	576
102	16	13121	14419	gi 149435	putative [Lactococcus lactis]	74	57	1299
108	4	4873	3902	gi 39478	ATP binding protein of transport ATPase [Bacillus firmus] ir[S15486]S15486 ATP-binding protein - Bacillus firmus pJP25946 YATR_BACFI HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN.	74	59	972
116	5	8574	7093	gi 1205430	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4342	4803	gi 146970	ribonucleoside triphosphate reductase [Escherichia coli] pir[A47331]A47331 anaerobic ribonucleotide reductase - Escherichia coli	74	58	462
121	7	5961	6581	gi 1107528	lsg start [Campylobacter coli]	74	51	621
128	3	2320	3531	gi 143318	phosphoglycerate kinase [Bacillus megaterium]	74	57	1212
130	7	5237	5791	gi 1256653	DNA-binding protein [Bacillus subtilis]	74	60	555
136	3	6745	5150	gi 143076	histidase [Bacillus subtilis]	74	58	1596
145	2	664	1368	gi 407773	deva gene product [Anabaena sp.]	74	45	705
152	1	552	277	gi 1377833	unknown [Bacillus subtilis]	74	54	276

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	saufq gene name	% sim	% ident	length (nt)
164	10	11064	11175	gi 580900	ORF3 gene product [Bacillus subtilis]	74	52	312
175	2	3109	2624	gi 642656	unknown [Rhizobium meliloti]	74	34	486
175	9	6064	5612	gi 854656	[Na/H antiporter system ORF2 [Bacillus alcalophilus]	74	46	453
195	11	11146	10339	gi 1204430	hypothetical protein (SP:P25745) [Haemophilus influenzae]	74	55	1008
205	17	9619	9059	gi 1044979	ribosomal protein L6 [Bacillus subtilis]	74	64	561
236	7	5574	6710	gi 1146207	[putative [Bacillus subtilis]	74	63	1117
241	3	4521	3334	gi 694121	[malate thiokinase [Methylobacterium extorquens]	74	52	1188
246	6	3305	2799	gi 467374	[single strand DNA binding protein [Bacillus subtilis]	74	64	507
249	4	6551	5313	gi 11524397	[glycine betaine transporter Opud [Bacillus subtilis]	74	55	1319
261	7	4389	4081	gi 809542	[CbrB protein [Erwinia chrysanthemi]	74	42	309
278	6	5714	4665	gi 1204872	[ATP-binding protein [Haemophilus influenzae]	74	54	1050
309	1	1220	666	gi 1205579	[hypothetical protein (G8:U14003.302) [Haemophilus influenzae]	74	53	555
315	2	1473	862	gi 143398	[quinol oxidase [Bacillus subtilis]	74	57	612
320	1	1	1065	gi 143389	[glutaminase of carbamyl phosphate synthetase [Bacillus subtilis] pif E39845 E39845 carbamoyl-phosphate synthase glutamine-hydrolyzing (EC 6.3.5.3), pyrimidine-repressible, small hain - Bacillus subtilis	74	60	1065
380	2	382	1128	gi 534857	[ATPase subunit a [Bacillus stearothermophilus]	74	56	747
405	2	1742	1311	gi 1303915	[yqhZ [Bacillus subtilis]	74	65	432
433	5	2503	3270	gi 473902	[alpha-acetolactate synthase [lactococcus lactis]	74	56	768
452	1	1	942	gi 413982	[ipa-5Br gene product [Bacillus subtilis]	74	57	942
461	1	3	1193	gi 558494	[homoserine dehydrogenase [Bacillus subtilis]	74	51	1193
461	2	1174	1407	gi 40211	[threonine synthase (thrC) (AA 1-352) [Bacillus subtilis] ir A25364 A25364 threonine synthase (EC 4.2.99.2) - Bacillus btills	74	56	234
462	2	402	734	gi 147520	[thioredoxin [Bacillus subtilis]	74	62	333
478	1	574	320	gi 1499005	[glycyl-tRNA synthetase [Methanococcus jannaschii]	74	52	255
501	2	739	1740	gi 217040	[acid glycoprotein [Streptococcus pyogenes]	74	58	1002
551	2	4083	2791	gi 143040	[glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pif D42728 D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC 4.3.8) - Bacillus subtilis	74	51	1293

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
573	1	1	477	gi1006605	hypothetical protein [Synecocystis sp.]	74	45	477
596	2	1780	1298	gi1303853	YggP [Bacillus subtilis]	74	55	483
618	2	2924	1758	gi1146237	21.4% of identity to trans-acting transcription factor of <i>Saccharomyces cerevisiae</i> ; 25% of identity to sucrose synthase of <i>Zea mays</i> ; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	gi1072380	[ORF] [Lactococcus lactis]	74	62	327
724	1	373	188	gi1143374	phosphoribosyl glycine synthetase (PUR-D; gtp start codon) <i>Bacillus subtilis</i>	74	58	186
743	2	604	1209	gi1153833	[ORF]; putative [Streptococcus parasanguis]	74	50	606
836	1	2	259	gi1143458	[ORF V] [Bacillus subtilis]	74	47	258
989	2	443	724	gi11303994	YqkM [Bacillus subtilis]	74	46	282
1106	1	1	492	gi1146970	lepD gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	gi11433948	lpa-24d gene product [Bacillus subtilis]	74	48	156
1234	1	817	472	gi1495245	recJ gene product [Erwinia chrysanthemi]	74	36	366
2586	1	2	238	gi11149701	sbcC gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	gi11405454	aconitase [Bacillus subtilis]	74	60	399
2962	1	650	363	gi1450586	3-phosphoglycerate kinase [Thermotoga maritima]	74	58	288
2983	1	3	191	gi11303893	YghL [Bacillus subtilis]	74	56	189
3018	1	2	223	gi1143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pirD42728 D42728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) - Bacillus subtilis	74	56	222
3038	1	510	256	pir S52915 S529	nitrate reductase alpha chain - <i>Bacillus subtilis</i> (fragment)	74	57	255
3062	1	374	189	gi1107528	ltg start [Campylobacter coli]	74	51	186
4035	1	184	360	gi1102725	unknown [Staphylococcus haemolyticus]	74	64	377
4045	1	607	305	gi11510977	M. jannaschii predicted coding region MJ0938 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	gi1520844	orf4 [Bacillus subtilis]	74	58	168
4449	1	3	221	gi1580910	peptide-synthetase ORF1 [Bacillus subtilis]	74	56	219
4587	1	458	231	gi11370207	orf6 [Lactobacillus sakei]	74	59	228

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi 146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pif[A29617/A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	74	60	186
4670	1	366	184	gi 256135	YbfB [Bacillus subtilis]	74	61	183
5	10	7953	7162	gi 143727	putative [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi 166338	dihydroxotetrate dehydrogenase [Agrobacterium tumefaciens]	73	55	1083
14	1	2024	1020	gi 143373	phosphoribosyl aminimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (Pur-H(II)) [Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi 1468939	meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae]	73	58	792
23	17	17379	16360	gi 297060	ornithine cyclodeaminase [Rhizobium meliloti]	73	37	1020
29	2	692	1273	gi 467442	stage V sporulation [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi 414000	ipa-76d gene product [Bacillus subtilis]	73	55	1554
37	8	8658	7402	gi 429259	pepr gene product [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi 168367	alpha-isopropylmalate isomerase (put-1; putative [Rhizomonas ircinellolides]	73	52	177
38	7	3931	4896	gi 405885	lysN [Escherichia coli]	73	58	966
44	6	5041	4238	gi 580895	unknown [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi 42909	moaB gene product [Escherichia coli]	73	50	540
45	3	2439	3080	gi 1109685	ProW [Bacillus subtilis]	73	47	642
54	13	14036	13794	gi 413931	ipa-7d gene product [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi 147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	73	53	819
65	1	1458	730	gi 677944	AppF [Bacillus subtilis]	73	56	729
80	2	1375	860	gi 580932	murD gene product [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi 580891	3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pif[A26522/A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis	73	55	1056
109	2	3493	2600	gi 1510849	M. jannaschii predicted coding region M0775 [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi 146970	ribonucleoside triphosphate reductase [Escherichia coli] pif[A47331/A47331 anaerobic ribonucleotide reductase - Escherichia coli	73	56	975
120	9	5726	6223	gi 1204333	anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]	73	62	498

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
132	5	4151	4363	gi1871048	HPSR2 - heavy chain potential motor protein [Giardia intestinalis]	73	43	213
140	6	5952	4324	gi1634107	kdpB [Escherichia coli]	73	59	1629
142	6	7060	5919	gi1410125	rfaH gene product [Bacillus subtilis]	73	57	1122
149	4	1866	1717	gi1460892	heparin binding protein-44, HBP-44 [mice, Peptide, 360 aa] pitJX0281JX0281 heparin-binding protein-44 precursor - mouse gi1220434 ORF [Mus musculus] (SUB 2-360)	73	53	150
158	1	1	1431	gi1882504	ORF_f560 [Escherichia coli]	73	57	1431
174	6	5152	4525	gi1146240	ketopentolate hydroxymethyltransferase [Bacillus subtilis]	73	55	828
175	8	5537	5178	gi1854657	Na/H antiporter system ORP3 [Bacillus alcalophilus]	73	56	360
186	5	6593	5493	gi1467477	unknown [Bacillus subtilis]	73	48	1101
249	6	6283	5729	gi11524397	glycine betaine transporter Opd [Bacillus subtilis]	73	56	555
265	4	1873	2280	gi139848	U3 [Bacillus subtilis]	73	41	408
270	1	328	582	gi1780461	220 kDa polyprotein (African swine fever virus)	73	53	255
278	4	4283	3618	gi11208965	hypothetical 23.3 kD protein [Escherichia coli]	73	49	666
279	3	4984	3593	gi1185288	isochorismate synthase [Bacillus subtilis]	73	58	1392
291	4	1207	1575	gi11511440	glutamine--fructose-6-phosphate transaminase [Methanococcus jannaschii]	73	63	369
299	2	735	1166	gi1467437	unknown [Bacillus subtilis]	73	58	432
299	5	2050	3234	gi1467439	temperature sensitive cell division [Bacillus subtilis]	73	53	1185
334	1	1237	728	gi1536655	ORF YBR244W [Saccharomyces cerevisiae]	73	43	510
336	2	1827	1036	gi1790943	urea amidolyase [Bacillus subtilis]	73	51	792
374	3	1389	1874	gi11405451	YneJ [Bacillus subtilis]	73	55	486
433	4	1916	2554	gi1473902	alpha-acetolactate synthase [Lactococcus lactis]	73	54	639
509	2	1795	1028	gi1467483	unknown [Bacillus subtilis]	73	56	768
513	1	1709	918	gi1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	73	56	792
533	2	239	733	gi11510605	hypothetical protein (SP:P42297) [Methanococcus jannaschii]	73	44	495
546	2	1148	2815	gi141748	hcdM protein (AA 1-520) [Escherichia coli]	73	52	1668
549	1	762	382	gi1114847	CinA [Bacillus subtilis]	73	57	381
567	1	1346	675	gi1410137	ORFX13 [Bacillus subtilis]	73	58	672

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
716	2	654	1112	gi1256623	exodeoxyribonuclease (Bacillus subtilis)	73	56	459
772	1	3	677	gi1142010	Shows 70.2% similarity and 49.6% identity to the EnvM protein of almonella cythamurium (Aaabeae sp.)	73	57	675
774	1	3	209	gi1409286	barU (Bacillus subtilis)	73	52	207
782	1	1	402	gi1143320	[gap] gene products (Bacillus megaterium)	73	56	402
789	2	451	762	gi11063246	low homology to p14 protein of Hemophilus influenzae and 14.2 kDa protein of Escherichia coli (Bacillus subtilis)	73	56	312
796	1	3	911	gi1853754	ABC transporter (Bacillus subtilis)	73	58	909
806	3	1209	949	gi1143786	cryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) p1r[JT0481]YMB5 tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus subtilis	73	51	261
816	2	4839	3097	gi141748	hcdM protein (AA 1-520) (Escherichia coli)	73	52	1743
839	1	798	400	gi1886906	argininosuccinate synthetase (Streptomyces clavuligerus) p1r[S57659]S57659 argininosuccinate synthetase (EC 6.3.4.5) - Streptomyces clavuligerus	73	59	399
857	1	3	290	gi1348052	acetoin utilization protein (Bacillus subtilis)	73	50	288
1008	1	790	398	gi140100	rodC (tag3) polypeptide (AA 1-746) (Bacillus subtilis) i1r[S06049]S06049 rodC protein - Bacillus subtilis p1r[13485]TAGF_BACSU TECHNIC ACID BIOSYNTHESIS PROTEIN F.	73	41	393
1018	1	1	213	gi1529357	No definition line found (Caenorhabditis elegans) sp1p46975[STT3_CAREL OLIGOSACCHARYL TRANSFERASE STT3] SUBUNIT OMLOG.	73	53	213
1013	1	3	491	gi1142706	comG1 gene product (Bacillus subtilis)	73	51	489
1174	1	395	204	gi1149513	alpha subunit of laminin 5 (Homo sapiens)	73	60	192
1175	1	655	329	gi1473817	"ORP" (Escherichia coli)	73	57	327
1187	1	3	209	gi1580870	lps-37D gene product (Bacillus subtilis)	73	52	207
1206	1	72	245	gi1144816	formyltetrahydrofolate synthetase (RhlFS) (ttg start codon) (EC 3.4.3) (Mooralla thermacetica)	73	43	174
1454	1	423	241	gi11213253	unknown (Schistosoma haematophyllum)	73	53	183
1469	1	517	260	gi11303787	YaeG (Bacillus subtilis)	73	55	258
1761	1	374	189	gi19135	Hst26Aa gene product (Drosophila simulans)	73	34	186
1849	1	467	243	gi1162307	DNA topoisomerase II (Trypanosoma cruzi)	73	60	225
2055	1	2	400	gi1559381	P47K protein (Rhodococcus erythropolis)	73	34	399
2556	1	2	244	gi1145925	feca (Escherichia coli)	73	62	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2947	2	549	400	gi1184680	polynucleotide phosphorylase [Bacillus subtilis]	73	51	150
2956	1	746	375	gi1143397	guinol oxidase [Bacillus subtilis]	73	58	372
3037	1	655	329	gi1143091	acetolactate synthase [Bacillus subtilis]	73	55	327
3115	1	385	194	gi1323866	overlapping out-of-phase protein [eggplant mosaic virus]	73	53	192
3603	2	700	527	gi1143521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	174
3743	1	798	400	gi1450688	hcdM gene of Ecopt1 gene product [Escherichia coli] pIR[S38437]S38437 hcdM protein - Escherichia coli pIR[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	73	54	399
3752	1	640	359	gi11524193	unknown [Mycobacterium tuberculosis]	73	59	282
3852	1	2	181	gi1216746	D-lactate dehydrogenase [Lactobacillus plantarum]	73	68	180
3914	1	475	239	gi1314901S134	hydroxymethylglutaryl-CoA synthase [EC 4.1.3.5] - Chicken (fragment)	73	53	237
3914	2	570	343	gi1528991	unknown [Bacillus subtilis]	73	38	228
4069	1	2	316	gi140003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] pIP3129[OD01_BACSU 2-OXOGLUTARATE DEHYDROGENASE: E1 COMPONENT [EC 2.4.2] (ALPHA- KETOGLUTARATE DEHYDROGENASE)]	73	55	315
4165	1	715	365	gi1143521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	351
4196	1	1	177	gi1409660	deoxyribose-phosphate aldolase [Bacillus subtilis] pIR[S09455]S09455 deoxyribose-phosphate aldolase [EC 4.1.2.4] - acillus subtilis	73	60	177
4202	1	572	378	gi1528991	unknown [Bacillus subtilis]	73	38	195
4314	1	2	193	gi1436797	N-acyl-L-amino acid amidohydrolase [Bacillus stearothermophilus] spIP37112]AMA_BACST N-ACYL-L-AMINO ACID AMIDOHYDROLASE [EC 5.1.1.14] (AMINOACYLASE)]	73	47	192
4393	1	3	263	gi1216267	ORF2 [Bacillus megaterium]	73	47	261
35	2	903	1973	gi1146196	phosphoglycerate dehydrogenase [Bacillus subtilis]	72	53	1071
38	22	19094	17877	gi1602031	similar to trimethylamine DH [Mycoplasma capricolum] pIR[S49950]S49950 probable trimethylamine dehydrogenase [EC 5.99.7] - Mycoplasma capricolum (SDC3) (fragment)	72	54	1218
38	23	18134	15162	gi1413968	ipa-64d gene product [Bacillus subtilis]	72	54	1029
44	19	11895	12953	gi1516272	unknown [Bacillus subtilis]	72	49	1059
48	7	6248	7117	gi143499	pyruvate synthase [Halobacterium halobium]	72	49	870
50	7	6563	5691	gi11205399	proton glutamate symport protein [Haemophilus influenzae]	72	53	873

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
53	9	10521	9259	gi 1303956	YojE (Bacillus subtilis)	72	52	1263
56	23	29549	29995	gi 467471	unknown (Bacillus subtilis)	72	47	447
69	4	5298	4123	gi 1356775	pfoS/R (Treponema pallidum)	72	46	1176
69	5	4377	4982	gi 904198	hypothetical protein (Bacillus subtilis)	72	43	606
73	1	2	856	gi 142997	glycerol uptake facilitator (Bacillus subtilis)	72	59	855
98	13	9371	10258	gi 467435	unknown (Bacillus subtilis)	72	50	888
127	1	1	1593	gi 217144	alanine carrier protein (thermophilic bacterium PS3)	72	56	1593
131	1	5197	2600	gi 153952	alanine transport protein - Thermophilic acterium PS-3	72	53	2598
141	4	1040	1978	gi 1405446	transketolase (Bacillus subtilis)	72	54	939
149	8	2819	2535	gi 606234	secY (Escherichia coli)	72	44	285
149	17	5472	5245	gi 1304472	DNA polymerase (Unidentified phycodnavirus clone OTU4)	72	55	228
154	1	1	210	gi 1205620	ferritin like protein (Haemophilus influenzae)	72	40	210
155	1	2207	1320	gi 391610	gamma-aminobutyrate synthase (Bacillus stearothermophilus)	72	57	888
180	1	2	328	gi 433630	pir JX0257 JX0257 geranyltransferase (EC 2.5.1.10) - Bacillus stearothermophilus	72	62	327
184	3	1145	3553	gi 1205110	virulence associated protein homolog (Haemophilus influenzae)	72	49	2409
195	2	1923	1279	gi 1001730	hypothetical protein (Synecocystis sp.)	72	45	645
206	13	14646	15869	gi 1064807	ORTHINE AMINOTRANSFERASE (Bacillus subtilis)	72	50	1224
209	2	462	932	gi 1204666	hypothetical protein (GB:K73124_53) (Haemophilus influenzae)	72	60	471
215	2	764	522	gi 881313	insulin receptor homolog (Drosophila melanogaster) pir S57245 S57245 insulin receptor homolog - fruit fly (Drosophila melanogaster) (SUB 46-2146)	72	63	243
224	1	2	790	gi 949974	luciferase repressor (Staphylococcus xylosus)	72	54	789
233	1	1526	765	gi 1408493	homologous to SwissProt:VIDA_EC001 hypothetical protein (Bacillus subtilis)	72	52	762
240	1	220	1685	gi 537049	ORF_0470 (Escherichia coli)	72	52	1266
245	1	3	1340	gi 1204578	hypothetical protein (GB:U06949_1) (Haemophilus influenzae)	72	46	1338

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
259	2	2108	1245	gi 1340128	ORF1 [Staphylococcus aureus]	72	59	864
304	2	285	1094	gi 1205330	glutamine-binding periplasmic protein [Haemophilus influenzae]	72	52	810
307	10	5326	5039	gi 1070015	protein-dependent [Bacillus subtilis]	72	53	288
315	1	517	260	gi 143399	quinol oxidase [Bacillus subtilis]	72	55	258
316	11	9622	9308	gi 1204445	hypothetical protein (SPI27857) [Haemophilus influenzae]	72	58	315
337	3	926	1609	gi 487433	citrate synthase II [Bacillus subtilis]	72	55	684
364	7	12538	10493	gi 1510643	ferrous iron transport protein B [Methanococcus jannaschii]	72	53	2046
409	2	340	1263	gi 1402944	orfRH1 gene product [Bacillus subtilis]	72	49	924
441	3	2177	1590	gi 1312379	highly conserved among eubacteria [Clostridium acetobutylicum]	72	48	588
453	6	2654	2505	pf 600601 BXSA	pf S34312 S34312 hypothetical protein V - Clostridium acetobutylicum	72	70	150
460	1	2	625	gi 1016162	antibacterial protein 3 - Staphylococcus haemolyticus	72	51	624
463	1	3253	1628	gi 666014	ABC transporter subunit [Cyanophora paradoxa]	72	60	1626
480	4	3047	3466	gi 433992	The polymorphism (RFLP) of this gene is associated with susceptibility to essential hypertension. The SA gene product has light homology to CoA synthetase [Homo sapiens]	72	53	420
502	1	1086	586	gi 310859	ATP synthase subunit epsilon [Bacillus subtilis]	72	50	501
519	1	81	1184	gi 1303704	ORF2 [Synecococcus sp.]	72	54	1104
559	1	3	746	gi 1107530	YrkE [Bacillus subtilis]	72	56	744
575	1	1142	573	gi 1303866	icaUD gene product [Campylobacter coli]	72	56	570
671	1	2	592	gi 1204497	Vqgs [Bacillus subtilis]	72	44	591
679	2	295	1251	gi 563758	protein-export membrane protein [Haemophilus influenzae]	72	52	957
687	2	295	957	gi 1146214	virulence-associated protein 2 [Dichelobacter nodosus]	72	49	663
837	1	1	435	gi 1146183	448 identical amino acids with the Escherichia coli ambA suppressor; putative [Bacillus subtilis]	72	54	435
868	1	150	788	gi 1377842	putative [Bacillus subtilis]	72	55	639
922	1	130	432	gi 1088269	unknown [Bacillus subtilis]	72	58	303
941	1	2	238	gi 153929	unknown protein [Azotobacter vinelandii]	72	49	237
980	1	840	421	gi 853767	NADPH-sulfite reductase flavoprotein component [Salmonella yphimurium]	72	59	420
					UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1209	1	383	213	gi 144735	neurotoxin type B (Clostridium botulinum)	72	44	171
1469	2	671	474	gi 1205458	hypothetical protein (CB.D26562.47) (Haemophilus influenzae)	72	63	198
1956	1	727	365	gi 154409	hexosephosphate transport protein (Salmonella typhimurium) p1r181853 81853 hexose phosphate transport system regulatory protein uhpB - Salmonella typhimurium	72	44	363
2101	1	3	401	gi 1303950	yqiY (Bacillus subtilis)	72	50	399
2503	1	569	399	gi 149713	formate dehydrogenase (Methanobacterium formicicum) p1r142712 A42712 formate dehydrogenase (EC 1.2.1.2) - ethanobacterium formicicum	72	56	171
2967	1	3	155	gi 1212729	yqjW (Bacillus subtilis)	72	46	153
3004	1	367	185	gi 665999	hypothetical protein (Bacillus subtilis)	72	55	183
3109	1	278	141	gi 413968	ipa-44d gene product (Bacillus subtilis)	72	45	138
3171	1	3	287	gi 515938	glutamate synthase (ferredoxin) (Synecocystis sp.) p1r1546957 546957 glutamate synthase (ferredoxin) (EC 1.4.1.1) - Synecocystis sp.	72	52	285
3771	1	26	367	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	72	63	342
3951	1	1	222	gi 1500409	M. jannaschii predicted coding region MJ1519 (Methanococcus jannaschii)	72	38	222
4190	1	721	362	gi 39956	11G1C (Bacillus subtilis)	72	57	360
4444	1	3	347	gi 1009366	respiratory nitrate reductase (Bacillus subtilis)	72	55	345
6	2	911	1200	gi 537095	ornithine carbamoyltransferase (Escherichia coli)	71	55	270
11	15	11350	10859	gi 532309	35 kDa protein (Escherichia coli)	71	47	492
19	2	1248	2435	gi 1244574	D-alanine:D-alanine ligase (Enterococcus hirae)	71	52	1188
21	2	898	1488	gi 149639	anthranilate synthase component 2 (Leptospira biflexa) p1r132840 C32840 anthranilate synthase (EC 4.1.3.27) component II Leptospira biflexa	71	45	591
34	1	1	567	gi 11303983	yqjF (Bacillus subtilis)	71	59	567
37	3	3192	2806	gi 1209681	glutamate-rich protein (Bacillus firmus)	71	50	387
38	18	12250	12462	gi 927645	arginyl endopeptidase (Porphyromonas gingivalis)	71	50	213
39	3	1246	4431	p1r1509411 S094	spoIIIE protein - Bacillus subtilis	71	49	3186
53	14	115770	14760	gi 142611	branched chain alpha-keto acid dehydrogenase E1-alpha (Bacillus subtilis)	71	58	1011
54	11	113461	12625	gi 141014	lont repressor (Bacillus subtilis)	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi 508175	ETIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	71	48	1293
57	18	13897	14334	gi 1063247	high homology to flavohemoprotein (Haemoglobin-like protein) of Alcaligenes eutrophus and Saccharomyces cerevisiae [Bacillus subtilis]	71	56	438
62	16	9831	10955	gi 1303926	YqjG [Bacillus subtilis]	71	54	1125
70	12	8505	8966	gi 147198	phnE protein [Escherichia coli]	71	38	462
86	5	2394	2089	gi 904205	hypothetical protein [Bacillus subtilis]	71	51	306
96	7	7601	8269	gi 709991	hypothetical protein [Bacillus subtilis]	71	49	669
100	6	4822	5931	gi 1060848	Opine dehydrogenase [Arthrobacter sp.]	71	45	1110
103	1	1062	532	gi 143089	lep protein [Bacillus subtilis]	71	41	531
109	18	15312	15695	gi 413985	lipa-6id gene product [Bacillus subtilis]	71	57	384
113	1	630	316	gi 663254	probable protein kinase [Saccharomyces cerevisiae]	71	57	315
114	5	6598	5603	gi 143156	membrane bound protein [Bacillus subtilis]	71	40	996
133	2	3087	1723	gi 1303913	YqhX [Bacillus subtilis]	71	53	1165
149	19	6335	5895	gi 529650	G40P [Bacteriophage SP21]	71	51	441
154	5	3635	3087	gi 425488	repressor protein [Streptococcus sobrinus]	71	47	549
164	11	11354	11689	gi 44318	ORF4 gene product [Bacillus subtilis]	71	52	136
169	5	1936	2745	gi 1403403	unknown [Mycobacterium tuberculosis]	71	56	810
193	2	272	1234	gi 1303788	YqeH [Bacillus subtilis]	71	49	963
205	1	1743	895	gi 1215694	GlnQ [Mycoplasma pneumoniae]	71	46	849
233	4	1849	2022	gi 633732	ORF1 [Campylobacter jejuni]	71	50	174
237	7	4501	5169	gi 149384	HistE [Lactococcus lactis]	71	54	669
272	4	2848	2273	gi 709993	hypothetical protein [Bacillus subtilis]	71	48	576
274	2	618	1496	gi 143035	NAD(P)H:glutamy]-transfer RNA reductase [Bacillus subtilis] pifA35252/A3522 5-aminolevulinate synthase (EC 2.3.1.37) - actillus subtilis	71	53	879
276	5	3349	2720	gi 303562	ORF210 [Escherichia coli]	71	50	630
287	1	136	660	gi 10634	20 kDa protein [Streptococcus gordonii]	71	53	525
288	6	3322	2771	gi 1256625	putative [Bacillus subtilis]	71	47	552

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
301	6	3492	2461	gi 467417	similar to lysine decarboxylase (Bacillus subtilis)	71	57	1032
306	4	6607	5222	gi 1256618	transport protein (Bacillus subtilis)	71	56	1386
307	2	1536	925	gi 632683	orfC (Mycoplasma capricolum)	71	45	612
310	5	5793	5146	gi 348052	acetoin utilization protein (Bacillus subtilis)	71	51	648
322	1	2	1303	gi 1001819	hypothetical protein (Synecocystis sp.)	71	46	1302
331	4	4171	3995	gi 467473	unknown (Bacillus subtilis)	71	57	177
350	2	548	922	gi 551879	ORF 1 (Lactococcus lactis)	71	55	375
375	4	1860	3071	gi 467447	unknown (Bacillus subtilis)	71	57	1212
380	5	1560	2102	gi 142557	ATP synthase b subunit (Bacillus megaterium)	71	43	543
414	2	251	637	gi 580904	homologous to E.coli rnpA (Bacillus subtilis)	71	49	387
424	1	335	1354	gi 581305	L-lactate dehydrogenase (Lactobacillus plantarum)	71	57	1020
436	4	3701	3270	pir P05031 P05	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)	71	66	432
482	1	3	1280	gi 410142	ORFX18 (Bacillus subtilis)	71	49	1278
525	3	2272	1844	gi 143370	phosphoribosylpyrophosphate amidotransferase (Pur-P; EC 2.4.2.14) Bacillus subtilis	71	56	429
529	4	2739	2047	gi 606150	ORF_309 (Escherichia coli)	71	43	693
563	1	22	969	gi 1237015	ORF4 (Bacillus subtilis)	71	53	948
581	1	506	255	gi 1301730	T2503.2 (Caenorhabditis elegans)	71	47	252
612	2	1068	913	gi 153968	fimbriae 2 (Salmonella typhimurium)	71	55	156
613	1	1	654	gi 466778	lysine specific permease (Escherichia coli)	71	50	654
618	1	1243	623	gi 1146238	poly(A) polymerase (Bacillus subtilis)	71	52	621
630	1	1170	586	gi 1486243	unknown (Bacillus subtilis)	71	53	585
691	1	1126	641	gi 289260	comE ORF1 (Bacillus subtilis)	71	51	486
694	2	149	427	gi 12971	NADH dehydrogenase subunit V (AA 1-605) (Gallus gallus) ir S10197 S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (SCC1)	71	47	279
715	2	169	777	gi 1303830	YqZ (Bacillus subtilis)	71	53	609
746	2	1473	970	gi 1377843	unknown (Bacillus subtilis)	71	52	504

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi1403459	YnaS [Bacillus subtilis]	71	49	636
753	1	1018	524	gi1510389	M. jannaachii predicted coding region MJ0296 [Methanococcus jannaachii]	71	53	495
761	1	3	215	gi1475972	pentafunctional enzyme (Pneumocystis carinii)	71	47	213
783	1	1203	703	gi1536655	ORF Y8R244w [Saccharomyces cerevisiae]	71	52	501
800	3	1292	987	gi1204326	tRNA delta(2)-isopentenylpyrophosphate transferase [Haemophilus influenzae]	71	48	306
806	1	116	286	gi11419075	cbiH gene product [Methanobacterium thermoautotrophicum]	71	50	171
931	1	973	488	gi1893358	PgsA [Bacillus subtilis]	71	56	486
1041	1	2	262	gi1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	71	45	261
1070	1	2	172	gi1709993	hypothetical protein [Bacillus subtilis]	71	46	171
1176	1	57	365	gi151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas nevalonii] p1c144756/A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	71	49	309
1181	1	366	184	gi146971	epiP gene product [Staphylococcus epidermidis]	71	50	183
1281	1	3	290	gi133016	ORF 419 protein [Staphylococcus aureus]	71	50	288
1348	1	456	229	gi1602683	orfC [Mycoplasma capricolum]	71	48	228
2002	1	756	379	gi11008177	ORF YJL046w [Saccharomyces cerevisiae]	71	48	378
2119	1	2	217	gi1046088	argDyl-tRNA synthetase [Mycoplasma genitalium]	71	50	216
2418	1	3	320	gi1499771	M. jannaachii predicted coding region MJ0936 [Methanococcus jannaachii]	71	57	318
2961	1	2	187	gi1312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	71	57	186
2999	2	67	306	gi1710020	nitrite reductase (nirB) [Bacillus subtilis]	71	43	240
3033	1	2	184	gi1262335	YnaA [Bacillus subtilis]	71	57	183
3584	1	3	338	gi1401716	beta-isopropylmalate dehydrogenase [Neurospora crassa]	71	55	336
3715	2	743	399	gi1563952	gluconate permease [Bacillus licheniformis]	71	59	345
3785	1	770	387	gi147382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	71	57	384
3875	1	541	272	gi1001541	hypothetical protein [Synchocystis sp.]	71	38	270
4135	1	637	320	gi142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	71	52	318
4249	1	63	239	gi1205363	deoxyribose aldolase [Haemophilus influenzae]	71	63	177
4508	1	530	267	gi1197667	vitellogenin [Anolis pulchellus]	71	46	264

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	3	1237	2721	gi 1321788	larginine ornithine antiporter [Clostridium perfringens]	70	54	1485
11	11	6572	7486	gi 216854	P47K [Pseudomonas chlororaphis]	70	41	915
12	1	2890	1481	gi 467230	replicative DNA helicase [Bacillus subtilis]	70	49	1410
15	1	1756	893	gi 451216	mannosephosphate isomerase [Streptococcus mutans]	70	46	864
15	2	1277	1050	gi 476092	unknown [Bacillus subtilis]	70	50	228
17	2	2112	1350	gi 165402	choline dehydrogenase [Escherichia coli]	70	52	783
21	1	2	925	gi 149516	anthranilate synthase alpha subunit [Lactococcus lactis] p1r[SJ5124] anthranilate synthase (EC 4.1.1.27) alpha chain - actococcus lactis subsp. lactis	70	50	924
25	7	5580	6251	gi 1389549	ORF3 [Bacillus subtilis]	70	52	672
33	6	6071	7423	gi 1303875	Yqha [Bacillus subtilis]	70	51	1353
36	2	959	1594	gi 500755	methyl purine glycosylase [Mus musculus]	70	47	636
38	8	4901	5860	gi 1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	70	44	960
44	8	5312	5989	gi 1006620	hypothetical protein [Synecocystis sp.]	70	49	678
46	10	8950	10020	gi 1403126	cszB gene product [Alcaligenes eutrophus]	70	45	1071
52	2	2727	1900	gi 1486247	unknown [Bacillus subtilis]	70	53	828
52	6	4048	4656	gi 244501	esterase II-carboxylesterase (EC 3.1.1.1) [Pseudomonas fluorescens, eptido, 218 aa]	70	50	609
56	8	8460	9962	gi 1339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	70	51	1503
62	1	48	290	gi 142702	A competence protein 2 [Bacillus subtilis]	70	47	243
64	1	1080	541	gi 1204377	acetylcholinesterase biosynthesis protein [Haemophilus influenzae]	70	47	540
70	5	5139	3595	gi 1204834	2',3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]	70	47	1545
91	4	7793	5466	gi 886471	methionine synthase [Catharanthus roseus]	70	56	2328
96	5	8754	7255	gi 1390961	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	70	54	1500
110	2	767	1300	gi 145294	adenine phosphoribosyl-transferase [Escherichia coli]	70	51	534
116	6	7026	7976	gi 143607	sporulation protein [Bacillus subtilis]	70	50	951
121	8	6401	6988	gi 1107528	ttg start [Campylobacter coli]	70	45	588
131	8	6842	7936	gi 1150454	protease PepQ [Lactobacillus delbrueckii]	70	48	1095

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
135	1	2	1489	gi 311309	putative membrane-bound protein with four times repetition of ro-ser-ala at the N-terminus; function unknown [Alcaligenes utrophus]	70	49	1488
138	3	418	714	gi 504181	hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	gi 49315	ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	15626	16618	gi 1205212	hypothetical protein (GB:010483_18) [Haemophilus influenzae]	70	50	993
205	2	2735	1803	gi 1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1386	gi 1204665	hypothetical protein (GB:X73124_26) [Haemophilus influenzae]	70	48	477
246	3	340	756	gi 215098	excisionase [Bacteriophage 154a]	70	46	417
263	7	7876	6749	gi 142540	aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	gi 1340128	ORF1 [Staphylococcus aureus]	70	50	906
302	6	3201	3827	gi 147782	ruvA protein (otg start) [Escherichia coli]	70	46	627
302	10	5879	7051	pir C38530 C385	queuine tRNA-ribosyltransferase (EC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1414	gi 1205934	aminopeptidase e/1 [Haemophilus influenzae]	70	46	1107
355	2	379	669	gi 1070013	protein-dependent [Bacillus subtilis]	70	48	291
401	1	1255	629	gi 733147	GuaF [Xanthomonas campestris]	70	33	627
444	10	8770	9273	gi 1204752	high affinity ribose transport protein [Haemophilus influenzae]	70	52	504
449	1	2	1243	gi 619724	HgtE [Bacillus firmus]	70	44	1242
472	1	637	320	gi 727145	open reading frame; putative [Bacillus mytiloliquefaciens] pir B29091 B29091 hypothetical protein (bgia region) - Bacillus mytiloliquefaciens (fragment)	70	41	318
480	2	727	1608	gi 142560	ATP synthase gamma subunit [Bacillus megaterium]	70	44	882
524	1	2	307	gi 602292	RCN2 protein [Brassica napus]	70	45	306
525	1	823	413	gi 143372	phosphoribosyl glycinate de formyltransferase (Pur-N) [Bacillus subtilis]	70	52	411
565	4	1625	2552	gi 881434	ORF1 [Bacillus subtilis]	70	51	1074
607	4	829	1284	gi 1511524	hypothetical protein (SP:P37002) [Methanococcus jannaschii]	70	50	456
633	1	1383	703	gi 431231	uracil permease [Bacillus caldolyticus]	70	53	681
646	3	1683	1309	gi 467340	unknown [Bacillus subtilis]	70	49	375
663	1	830	417	gi 1303873	Y-02 [Bacillus subtilis]	70	40	414

TABLE 2

2. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
681	1	1488	781	gi 1001678	hypothetical protein (Synecocystis sp.)	70	53	708
708	1	2	448	sp P33940 YQJH_	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-AUKB INTERGENIC REGION.	70	51	447
725	1	51	722	gi 1001644	hypothetical p. tein (Synecocystis sp.)	70	46	672
776	1	1371	787	gi 145165	putative (Escherichia coli)	70	47	585
834	1	250	783	gi 552971	NADH dehydrogenase (ndhF) (Vicia faba)	70	47	534
865	2	1585	1379	gi 1204636	ATP-dependent helicase (Haemophilus influenzae)	70	45	207
894	1	535	269	gi 467364	DNA binding protein (probae) (Bacillus subtilis)	70	41	267
919	1	3	317	gi 1314847	CinA (Bacillus subtilis)	70	40	315
944	1	3	572	gi 709991	hypothetical protein (Bacillus subtilis)	70	44	570
988	2	772	605	gi 132441	ORF 3: putative (Bacillus subtilis)	70	50	168
1055	1	3	335	gi 529755	ispeC (Streptococcus pyogenes)	70	37	333
1093	1	2	904	gi 851754	ABC transporter (Bacillus subtilis)	70	49	903
1109	1	2	310	gi 1001827	hypothetical protein (Synecocystis sp.)	70	42	309
1220	1	468	235	pir S23616 S236	lepiB protein - Staphylococcus epidermidis	70	40	234
1279	1	73	348	gi 153015	FenA protein (Staphylococcus aureus)	70	47	276
1336	1	195	542	sp P31776 PBPA_	PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (PENICILLIN-BINDING PROTEIN A).	70	50	348
1537	2	232	402	gi 1146181	putative (Bacillus subtilis)	70	50	171
1574	1	451	272	gi 219630	endothelin-A receptor (Homo sapiens)	70	47	180
1640	1	690	346	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)	70	46	345
2504	1	2	286	gi 495179	transmembrane protein (Lactococcus lactis)	70	51	285
3061	1	564	301	gi 508175	EFIC domain of PTS-dependent Gal transport and phosphorylation Escherichia coli	70	44	264
3128	1	2	199	gi 1340096	unknown (Mycobacterium tuberculosis)	70	51	198
3218	1	3	488	gi 515938	glutamate synthase (ferredoxin) (Synecocystis sp.) pir S46957 S46957	70	50	486
3323	1	794	399	gi 1154891	Glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	70	52	396
3679	1	599	399	gi 529385	ATP binding protein (Phormidium laminosum)	70	30	201
					chromosome condensation protein (Caenorhabditis elegans)			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3841	1	706	398	gi11208965	hypothetical 23.3 kd protein [Escherichia coli]	70	47	309
3929	1	3	401	gi1149435	putative [Lactococcus lactis]	70	49	399
4044	1	595	374	gi1602031	similar to trimethylamine DH [Mycoplasma capricolum] pIR[S49950]S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	70	40	222
4329	1	558	280	gi11339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	70	49	279
4422	1	576	289	gi1296464	ATPase [Lactococcus lactis]	70	57	288
4647	1	361	200	gi1166412	NADH-glutamate synthase [Medicago sativa]	70	59	162
16	8	7571	9031	gi11499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	69	44	1461
16	9	9080	10033	gi11353197	thioredoxin reductase [Eubacterium acidaminophilum]	69	54	934
30	1	1452	727	gi11204910	hypothetical protein (G8:U14003_302) [Haemophilus influenzae]	69	52	726
38	4	1023	1298	gi1407773	devA gene product [Anabaena sp.]	69	41	276
44	9	5987	6595	gi11205920	molybdate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]	69	45	609
62	15	9104	9475	gi1385178	unknown [Bacillus subtilis]	69	44	372
66	4	2402	2803	gi11303893	YqkL [Bacillus subtilis]	69	51	402
67	15	14124	13627	gi1149647	ORF2 [Listeria monocytogenes]	69	37	498
67	17	14053	14382	gi1305002	ORF.f356 [Escherichia coli]	69	49	330
67	19	15130	15807	gi1109684	ProV [Bacillus subtilis]	69	45	678
78	3	1447	2124	gi11256633	putative [Bacillus subtilis]	69	53	678
78	4	4513	3725	gi11103958	YqjG [Bacillus subtilis]	69	32	789
85	4	4521	4213	pir129326 E293	hypothetical protein (pur operon) - Bacillus subtilis	69	32	309
86	6	3253	2654	gi1073332	OrfC [Bacillus subtilis]	69	50	600
95	1	96	710	gi178646A	4A11 antigen, sperm cell membrane antigen-putative sucrose-specific phosphotransferase enzyme II homolog [mice, testis, Peptide Partial, 72 aa]	69	43	615
100	7	6023	7426	gi11205355	Na ⁺ /H ⁺ antiporter [Haemophilus influenzae]	69	39	1404
102	2	2678	1650	gi1561690	isialoglycoprotease [Pasteurella haemolytica]	69	47	1029
103	8	11241	8537	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	69	54	3705
103	11	14987	12552	gi1710020	nitrite reductase (nirB) [Bacillus subtilis]	69	51	2436

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi 154411	hexosephosphate transport protein [Salmonella typhimurium] p1r p41853 041853 hexose phosphate transport system protein uhpT - salmonella typhimurium	69	51	1461
112	16	18644	17414	gi 1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	69	50	721
113	2	33	953	gi 290509	lo307 [Escherichia coli]	69	43	921
114	2	1537	1058	gi A42771 A427	reticulocyte-binding protein 1 - Plasmodium vivax	69	39	480
121	6	4309	5310	gi 1154633	WdF [Bacillus subtilis]	69	53	1002
125	2	267	854	gi 413931	lpa-7d gene product [Bacillus subtilis]	69	43	588
149	27	10666	10400	gi S28089 S280	hypothetical protein A - yeast (Zygosaccharomyces bisporus) plasmid pSD	69	39	267
161	1	1598	813	gi 1205538	hypothetical protein (G8:U14003_302) [Haemophilus influenzae]	69	47	786
165	4	2222	4633	gi 140054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus subtilis]	69	52	2412
169	3	1210	1761	gi 296031	elongation factor Ts [Spirulina platensis]	69	45	552
175	12	8686	8339	gi 732882	Pine protein [Escherichia coli]	69	69	348
190	2	484	1671	gi P17731 H158	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)	69	48	1188
206	1	5551	2777	gi 41750	hadr protein (AA 1-1033) [Escherichia coli]	69	49	2775
206	4	6038	5796	gi 1256135	Ybhp [Bacillus subtilis]	69	48	243
249	1	636	319	gi 1404456	YnuP [Bacillus subtilis]	69	40	110
302	8	4820	5776	gi 1001768	hypothetical protein [Synechocystis sp.]	69	48	957
324	2	7384	3893	gi 1256798	pyruvate carboxylase [Rhizobium etli]	69	53	3492
351	3	2098	1808	gi 1491664	T04H1.4 [Caenorhabditis elegans]	69	30	291
369	3	2075	2105	gi 336458	ORP [Balaenoptera acutorostrata]	69	61	231
392	3	1999	2424	gi 556015	ORP1 [Bacillus subtilis]	69	45	426
410	1	87	779	gi 155611	phosphoglyceromutase [Zymomonas mobilis]	69	58	693
421	1	2085	1129	gi 1276985	arginase [Bacillus caldovelox]	69	54	957
448	8	6713	7741	gi 1221782	purine synthesis repressor [Haemophilus influenzae]	69	40	1029
453	1	828	415	gi 1122758	unknown [Bacillus subtilis]	69	57	414
469	2	3286	2246	gi 1458228	mutY homolog [Homo sapiens]	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
509	3	1730	1371	gi 49224	ORF 4 [Synchococcus sp.]	69	39	360
520	5	3023	2823	gi 726427	similar to D. melanogaster MST101-2 protein (PIR:S34134) Caenorhabditis elegans	69	39	201
531	1	26	760	gi 509672	repressor protein [Bacteriophage Tuc2009]	69	33	735
589	1	107	253	gi 169101	[17.9 kDa heat shock protein (hsp17.9) [Pisum sativum]	69	52	147
594	2	597	1191	gi 142783	DNA photolyase [Bacillus firmus]	69	48	795
604	4	2476	2114	gi 413930	[ipa-6d gene product [Bacillus subtilis]	69	45	363
607	1	2	313	gi 1236103	[W08D2.3 [Caenorhabditis elegans]	69	47	312
607	2	590	312	gi 536715	[ORF YBR275c [Saccharomyces cerevisiae]	69	39	279
734	1	864	433	gi 467327	unknown [Bacillus subtilis]	69	44	432
759	1	3	338	gi 1009367	Respiratory nitrate reductase [Bacillus subtilis]	69	50	336
761	2	392	586	gi 13508	[Leucyl-tRNA synthetase [Cycloplasmic] [Saccharomyces cerevisiae] [1370340 ORF YPL160w [Saccharomyces cerevisiae]	69	46	195
802	1	72	1013	gi 143044	[ferrochelatase [Bacillus subtilis]	69	55	942
816	1	2573	1368	gi 1510268	[restriction modification system S subunit [Methanococcus jannaschii]	69	45	1206
838	2	133	387	gi 1255371	[coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.1; Similar to guanylate kinase [Caenorhabditis elegans]	69	46	255
851	2	745	1005	gi 288998	[scA gene product [Antilhammon sp.]	69	39	261
867	1	535	269	gi 1070014	[protein-dependent [Bacillus subtilis]	69	47	267
995	1	954	478	gi 1205569	[transcription elongation factor [Haemophilus influenzae]	69	53	477
999	1	1009	506	gi 89256	[predicted trithorax protein [Drosophila virilis]	69	21	504
1127	1	1315	659	gi 1205434	[H. influenzae predicted coding region H1191 [Haemophilus influenzae]	69	56	657
1138	1	248	460	gi 1510646	[M. jannaschii predicted coding region M20568 [Methanococcus jannaschii]	69	48	213
2928	1	3	401	gi 290503	[glutamate permease [Escherichia coli]	69	41	399
3090	1	444	223	gi 1204987	[DNA polymerase III, alpha chain [Haemophilus influenzae]	69	36	222
3817	1	2	400	gi 1483199	[peptide-synthetase [Amycolatopsis mediterranei]	69	45	399
3833	1	667	335	gi 1524193	[unknown [Mycobacterium tuberculosis]	69	46	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi 546918	orfY 3' of comK [Bacillus subtilis, E26, peptide Partial, 140 aa] pir543612 S43612 hypothetical protein Y - Bacillus subtilis sp P40398 YHKO_BACSU HYPOTHETICAL PROTEIN IN COMK 3' REGION (ORFY FRAGMENT)	69	64	348
4115	2	215	400	gi 517205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	69	59	186
4139	1	1	333	gi 1208451	hypothetical protein [Synchocytis sp.]	69	36	333
4258	1	457	230	gi 496158	restriction-modification enzyme subunit M1 [Mycoplasma pulmonis] pir S49395 S49395 HsdM1 protein - Mycoplasma pulmonis [SGC3]	69	43	228
4317	1	90	374	gi 413967	ipa-43d gene product [Bacillus subtilis]	69	44	285
4465	1	3	293	gi 396296	similar to phosphotransferase system enzyme II [Escherichia coli] sp P32672 PTMC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT PHOSPHOTRANSFERASE ENZYME II, C COMPONENT	69	49	291
3	1	2302	1193	gi 1109685	ProW [Bacillus subtilis]	68	46	1110
15	4	2592	2074	gi 807973	unknown [Saccharomyces cerevisiae]	68	45	519
31	8	6328	8772	gi 290642	ATPase [Enterococcus hirae]	68	48	2445
40	2	1115	750	gi 606342	ORF_0622, reading frame open far upstream of start; possible rameshift, linking to previous ORF [Escherichia coli]	68	55	366
46	9	6886	8415	gi 355276	aldehyde dehydrogenase [Vibrio cholerae]	68	44	1530
48	3	3643	3404	gi 285608	241k polypeptide [Apple stem grooving virus]	68	47	240
48	4	3536	4132	gi 1045937	H. genitalium predicted coding region HQ246 [Mycoplasma genitalium]	68	39	597
53	10	11671	10685	gi 1303952	[YqJA [Bacillus subtilis]	68	46	987
70	9	7346	8155	gi 147198	phnE protein [Escherichia coli]	68	40	810
89	4	1899	2566	gi 145173	35 kDa protein [Escherichia coli]	68	43	1068
108	1	2187	1150	gi 38722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir A29277 A29277 aldose 1-epimerase [EC 5.1.3.3] - Acinetobacter calcoaceticus	68	57	1036
112	5	2666	3622	gi 133724	MalC [Streptococcus pneumoniae]	68	55	957
116	7	7865	8638	gi 143608	sporulation protein [Bacillus subtilis]	68	48	774
118	3	2484	3698	gi 1103805	YqeR [Bacillus subtilis]	68	46	1215
120	2	1424	1594	sp P36038 CYSU_	SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT [EC 1.8.1.2] (SIR-FP)	68	45	171
129	1	1	1011	gi 396307	argininosuccinate lyase [Escherichia coli]	68	50	1011

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	Length (nt)
132	3	1867	2739	gi 16267	ORF2 (Bacillus megaterium)	68	48	873
134	2	848	1012	gi 147545	DNA recombinase (Escherichia coli)	68	50	165
141	2	372	614	gi 872116	act (stress inducible protein) (Glycine max)	68	48	243
149	7	2454	2260	gi 145774	hsp70 protein (dnaK gene) (Escherichia coli)	68	36	243
155	2	1776	1534	gi 216583	ORF1 (Escherichia coli)	68	51	1464
158	3	1826	3289	sp P33940 YOUN	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION	68	46	570
169	6	2749	3318	gi 1403402	unknown (Mycobacterium tuberculosis)	68	51	1794
175	10	9158	7165	gi 1072395	phaA gene product (Rhizobium meliloti)	68	48	1251
188	7	4184	5434	gi 1173843	3-ketoacyl-ACP synthase II (Vibrio harveyi)	68	55	759
189	3	907	1665	gi 467383	DNA binding protein (probable) (Bacillus subtilis)	68	48	975
206	5	7683	6709	gi 1256138	YbbI (Bacillus subtilis)	68	48	1752
206	8	10425	12176	gi 452687	pyruvate decarboxylase (Saccharomyces cerevisiae)	68	39	228
212	8	3421	3648	gi 1369941	cl gene product (Bacteriophage B1)	68	45	1026
214	8	5457	6482	gi 1420467	ORF YOR196c (Saccharomyces cerevisiae)	68	46	582
237	4	2507	3088	gi 149381	Hish (Lactococcus lactis)	68	47	999
243	5	5540	4542	gi 1235684	mevalonate pyrophosphate decarboxylase (Saccharomyces cerevisiae)	68	42	162
262	1	3	164	gi 150974	4-oxalocrotonate tautomerase (Pseudomonas putida)	68	49	867
262	2	1984	1118	gi 1147744	PSR (Enterococcus hirae)	68	50	564
276	6	3702	3139	sp P30750 ABC_E	ATP-BINDING PROTEIN ABC (FRAGMENT)	68	53	621
306	6	6345	5725	gi 1256617	adenine phosphoribosyltransferase (Bacillus subtilis)	68	45	750
333	3	4599	3850	gi 467473	unknown (Bacillus subtilis)	68	45	180
365	6	5017	4838	gi 1130643	T2283.3 (Caenorhabditis elegans)	68	51	1098
376	2	549	1646	gi 1277026	DAPA aminotransferase (Bacillus subtilis)	68	47	870
405	1	1741	872	gi 1303917	YqjB (Bacillus subtilis)	68	44	315
406	2	853	539	gi 1511513	ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)	68	48	168
426	6	3558	3391	gi 624632	GltL (Escherichia coli)	68	43	222
438	1	108	329	gi 146923	nitrogenase reductase (Escherichia coli)	68	43	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
443	1	476	240	gi 535810	hlppuricase [Campylobacter jejuni]	68	42	237
443	2	518	1015	gi 1204742	H. influenzae predicted coding region H10491 (Haemophilus influenzae)	68	48	498
443	5	4447	3779	gi f06660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir 549455 549455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	68	55	669
476	2	240	1184	gi 571345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 YWIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION.	68	45	945
486	2	1876	1046	gi 147128	transport protein [Escherichia coli]	68	41	831
517	3	1764	2084	gi 523809	orf2 [Bacteriophage A2]	68	64	321
572	1	2	571	sp P39237 Y05L	HYPOTHETICAL 6.8 KD PROTEIN IN NRDC-TK INTERGENIC REGION.	68	47	570
646	1	914	459	gi 413982	ipa-58r gene product [Bacillus subtilis]	68	52	456
659	3	1668	1901	gi 1407541	C33D9.8 [Caenorhabditis elegans]	68	36	234
864	5	1510	1716	gi 145774	hap70 protein (dnaK gene) [Escherichia coli]	68	48	207
920	1	860	432	gi 1510416	hypothetical protein (SP-P1466) [Methanococcus jannaschii]	68	54	429
952	1	1096	611	gi C33456	reductase [Leishmania major]	68	46	486
970	1	91	402	gi 1354775	pfoS/R [Treponema pallidum]	68	46	312
1028	1	1064	534	gi C40117	diaminopimelate decarboxylase [Bacillus subtilis]	68	47	531
1029	1	438	216	gi 1335714	Plasmidium falciptarum mRNA for asparagine-rich antigen (clone 17C1) [Plasmidium falciptarum]	68	31	213
1058	1	692	348	gi 501649	epic gene product [Staphylococcus epidermidis]	68	46	345
1096	2	665	465	gi 143434	Rho Factor [Bacillus subtilis]	68	43	201
1308	1	2	694	gi 169939	group B oligopeptidase, PepB [Streptococcus agalactiae]	68	50	693
1679	1	2	238	gi 17205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	68	53	237
2019	1	3	383	gi 153898	transport protein [Salmonella typhimurium]	68	51	381
2077	1	3	326	pir C33496 C334	hisC homolog - Bacillus subtilis	68	47	324
2112	1	613	374	gi 164884	lamin LII [Xenopus laevis]	68	50	240
2273	1	793	398	gi 581648	epiB gene product [Staphylococcus epidermidis]	68	45	396
2348	1	2	385	gi 216869	branched-chain amino acid transport carrier [Pseudomonas aeruginosa] p C138534 A38534 branched-chain amino acid transport protein braz [Pseudomonas aeruginosa]	68	41	384

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2955	1	768	400	gi 904179	hypothetical protein [Bacillus subtilis]	68	49	369
2981	1	572	288	gi 508979	GTP-binding protein [Bacillus subtilis]	68	48	285
3014	1	584	294	gi 1524394	ORF-2 upstream of gbaB operon [Bacillus subtilis]	68	45	291
3082	1	336	169	gi 1204696	fructose-permease IIBC component [Haemophilus influenzae]	68	53	168
3108	1	103	258	gi 217855	heat-shock protein [Arabidopsis thaliana]	68	48	156
3639	1	919	461	gi 1510490	nitrate transport permease protein [Methanococcus jannaschii]	68	47	459
3657	1	1	330	gi 155369	PTS enzyme-II fructose [Xanthomonas campestris]	68	48	330
3823	1	780	391	gi 603768	HutU protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	68	54	390
3982	1	2	277	gi 149435	putative [Lactococcus lactis]	68	47	276
4051	1	1	342	gi 450688	hsdM gene of EcoPrf1 gene product [Escherichia coli] pir[S38437/S38437 hsdM protein - Escherichia coli pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	48	342
4089	1	12	209	gi 1353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	68	47	198
4143	1	47	187	gi 603769	HutU protein, urocanase [Bacillus subtilis]	68	55	141
4148	1	2	352	gi 450688	hsdM gene of EcoPrf1 gene product [Escherichia coli] pir[S38437/S38437 hsdM protein - Escherichia coli pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	51	351
4173	1	2	382	gi 1041097	Pyruvate Kinase [Bacillus psychrophilus]	68	48	381
4182	1	498	250	gi 413968	lipa-44d gene product [Bacillus subtilis]	68	50	249
4362	2	168	318	gi 450688	hsdM gene of EcoPrf1 gene product [Escherichia coli] pir[S38437/S38437 hsdM protein - Escherichia coli pir[S09629/S09629 hypothetical protein A - Escherichia coli] (SUB 40-520)	68	44	171
5	11	9493	8300	gi 143727	putative [Bacillus subtilis]	67	46	1194
31	11	10318	9833	gi 216746	l0-lactate dehydrogenase [Lactobacillus plantarum]	67	41	486
32	3	1560	3155	gi 1098557	renal sodium/dicarboxylate cotransporter [Homo sapiens]	67	46	1596
32	5	4945	4145	gi 1510720	prephenate dehydratase [Methanococcus jannaschii]	67	51	801
36	5	5350	4268	gi 146216	45% identity with the product of the ORF6 gene from the Erwinia herbicola carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5304	gi 1006621	hypothetical protein [Synecocystis sp.]	67	43	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	7	3943	8481	gi 304131	glutamate synthase large subunit precursor (Azospirillum brasilense) pIR1846602 846602 glutamate synthase (NAADPH) (EC 1.4.1.13) alpha chain - Azospirillum brasilense	67	52	4539
56	12	13923	14678	gi 1000453	TreR (Bacillus subtilis)	67	48	756
62	8	5092	4757	gi 1113949	orf3 (Bacillus, C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa)	67	45	336
62	10	7570	6338	gi 654655	Na/H antiporter system (Bacillus alcalophilus)	67	49	1233
99	3	2119	3321	gi 1204349	hypothetical protein (GB1GBD9012_3) (Haemophilus influenzae)	67	50	1203
102	9	5695	7176	gi 149432	putative (Lactococcus lactis)	67	51	1482
103	13	14549	14049	gi 1408497	LPD gene product (Bacillus subtilis)	67	48	501
109	15	14821	13982	gi 413976	lipo-32r gene product (Bacillus subtilis)	67	49	840
109	17	14811	15194	gi 413983	lipo-59d gene product (Bacillus subtilis)	67	29	384
121	4	1713	2153	gi 1262335	YmaA (Bacillus subtilis)	67	54	441
122	1	1	1149	gi 143047	ORF8 (Bacillus subtilis)	67	35	1149
124	5	4060	3518	gi 556885	unknown (Bacillus subtilis)	67	47	543
131	2	4584	3589	gi 1046081	hypothetical protein (GB1D26185_10) (Mycoplasma genitalium)	67	30	996
140	3	2899	2297	gi 146549	kdpC (Escherichia coli)	67	45	603
142	4	5409	4198	gi 1212775	GTP cyclohydrolase II (Bacillus anisotrophicus)	67	55	1212
147	5	2913	2374	gi 1303709	YrkA (Bacillus subtilis)	67	44	540
152	8	6341	6673	gi 1377841	unknown (Bacillus subtilis)	67	48	333
161	4	2720	3763	gi 496319	SphA (Synechococcus sp.)	67	47	1044
163	6	1989	3428	gi 595681	2-oxoglutarate/malate translocator (Spinacia oleracea)	67	47	1440
193	3	1351	1626	gi 1511101	shikimate 5-dehydrogenase (Methanococcus jannaschii)	67	53	276
200	2	917	2179	gi 142439	ATP-dependent nuclease (Bacillus subtilis)	67	48	1263
206	10	12445	12801	sp P37347 YECD_	HYPOTHEICAL 21.8 KD PROTEIN IN ASPS 5' REGION.	67	47	357
206	11	13047	14432	gi 732813	branched-chain amino acid carrier (Lactobacillus delbrueckii)	67	46	1386
208	2	1321	809	gi 1033037	100 kDa heat shock protein (Hsp100) (Leishmania major)	67	36	513
238	3	1039	2052	gi 809542	CbrB protein (Erwinia chrysanthami)	67	42	1014

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
246	2	176	367	gi 315098	excisionase [Bacteriophage 154a]	67	37	192
276	2	2260	1412	gi 303560	ORF271 [Escherichia coli]	67	50	849
297	6	2223	3056	gi 142784	CtaA protein [Bacillus firmus]	67	46	834
307	7	5220	4186	gi 1070013	protein-dependent [Bacillus subtilis]	67	43	1035
316	1	36	1028	gi 1161061	dioxygenase [Methylobacterium extorquens]	67	52	993
324	3	5650	5030	gi 1469784	putative cell division protein ftaW [Enterococcus hirae]	67	49	621
336	1	524	264	gi 173122	urea amidolyase [Saccharomyces cerevisiae]	67	45	261
340	1	108	1194	cp P30053 SYHLS	[HISTIDYL-TRNA SYNTHETASE [EC 6.1.1.21] (HISTIDINE--TRNA LIGASE) (HISRS5).	67	47	1287
364	3	4890	3592	gi 151259	HMGR-CoA reductase [EC 1.1.1.88] [Pseudomonas mavalonii] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.	67	46	1299
365	3	2940	2113	gi 1296823	orf2 gene product [Lactobacillus helveticus]	67	47	828
367	2	325	918	gi 1039479	ORFU [Lactococcus lactis]	67	47	594
395	3	666	1271	gi 1204516	hypothetical protein (GB:U00014.4) [Haemophilus influenzae]	67	55	606
415	1	1800	901	gi 282579	CG Site No. 29739 [Escherichia coli]	67	46	900
419	1	1799	903	gi 520752	putative [Bacillus subtilis]	67	48	897
474	1	2	796	gi A86906	argininosuccinate synthetase [Streptomyces clavuligerus] pir S57659 S57659 argininosuccinate synthase [EC 6.3.4.5] - treptonmyces clavuligerus	67	49	795
485	2	1921	2226	gi 143434	Rho Factor [Bacillus subtilis]	67	43	306
596	1	1728	865	gi 1303853	YggF [Bacillus subtilis]	67	47	864
700	1	433	218	gi 1204628	hypothetical protein (SP:P21498) [Haemophilus influenzae]	67	47	216
806	2	249	647	gi 677947	AppC [Bacillus subtilis]	67	51	399
828	2	340	900	gi 777761	lrrA [Synechococcus sp.]	67	37	561
833	1	1407	916	gi 142996	regulatory protein [Bacillus subtilis]	67	41	492
856	1	1555	779	gi 780224	ZK970.2 [Caenorhabditis elegans]	67	38	777
888	1	1614	850	gi 437315	77C start codon [Bacillus licheniformis]	67	40	765
1034	1	1190	597	gi 1205113	hypothetical protein (GB:L19201.15) [Haemophilus influenzae]	67	45	594
1062	1	636	319	gi 1303850	YggC [Bacillus subtilis]	67	41	318
1067	1	918	460	ptr A32950 A329	probable reductase protein - Leishmania major	67	54	459

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1358	1	3	293	gi11001369	hypothetical protein [Synecococcus sp.]	67	44	291
2381	1	3	302	gi11510416	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	67	48	300
3000	1	1	507	gi11517205	[67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogones]	67	56	507
3066	1	464	234	gi1308861	ORF start codon [Lactococcus lactis]	67	46	231
3087	1	454	251	gi11205366	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	67	44	204
3101	1	2	256	gi11531541	uroporphyrinogen III methyltransferase [Zea mays]	67	55	255
3598	1	728	393	gi1151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas nevalonii] pir[A44756]A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	56	336
3765	2	584	366	gi11557489	menD [Bacillus subtilis]	67	45	219
3788	1	658	398	pir[S22915]S229	nitrate reductase alpha chain - Bacillus subtilis (fragment)	67	45	261
3883	1	2	265	gi1704397	cystathionine beta-lyase [Arabidopsis thaliana]	67	46	264
3926	1	2	340	gi11483199	peptide-synthetase [Amycolatopsis mediterranea]	67	44	339
4417	1	82	396	gi11205337	ribonucleotide transport ATP-binding protein [Haemophilus influenzae]	67	46	315
2	3	3075	1989	gi11555348	CodV [Bacillus subtilis]	66	42	915
15	6	2273	2542	gi1146491	SmtB [Synecococcus PCC7942]	66	37	270
31	9	8059	7826	gi1292046	mucin [Homo sapiens]	66	44	234
31	10	9034	9258	gi11204545	mercury scavenger protein [Haemophilus influenzae]	66	48	224
32	6	6347	5253	gi1998342	inducible nitric oxide synthase [Gallus gallus]	66	47	1095
44	13	8856	10124	gi11510751	molybdenum cofactor biosynthesis mcoA protein [Methanococcus jannaschii]	66	46	1289
48	2	1276	2868	gi1150209	ORF 1 [Mycoplasma mycoides]	66	40	1593
58	8	7178	8428	gi1665999	hypothetical protein [Bacillus subtilis]	66	47	1251
62	7	5143	4370	gi11072398	phd gene product [Rhizobium meliloti]	66	40	774
70	14	11693	10998	gi1809660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir[S49455]S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	66	55	696
76	1	1	1305	gi1142440	ATP-dependent nuclease [Bacillus subtilis]	66	42	1305
91	6	9236	8205	gi1704397	cystathionine beta-lyase [Arabidopsis thaliana]	66	43	1032
102	5	3810	3265	gi11204323	hypothetical protein (SP:P31805) [Haemophilus influenzae]	66	41	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Accession	Match name	% sim	% ident	length (nt)
103	4	3418	2732	gi 971344		nitrate reductase gamma subunit [Bacillus subtilis] ap P42177 NARI_BACSU NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4) . gi 1009369 Respiratory nitrate reductase [Bacillus subtilis] (SUB -160)	66	48	697
109	6	4243	4674	gi 170886		glucosamine-6-phosphate deaminase [Candida albicans] pir A46652 A46652 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - east [Candida albicans]	66	45	432
112	17	17491	17712	gi 1323179		ORF YGR11W [Saccharomyces cerevisiae]	66	33	222
116	2	4667	2637	gi 1491813		gamma-glutamyltranspeptidase [Bacillus subtilis]	66	43	2031
150	5	3189	2989	gi 1146224		putative [Bacillus subtilis]	66	30	201
172	5	3264	3662	gi 755152		highly hydrophobic integral membrane protein [Bacillus subtilis] ap P42953 TAGC_BACSU TEICHOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGG.	66	41	399
174	5	4592	3723	gi 1146241		pantothenate synthetase [Bacillus subtilis]	66	49	870
175	4	3209	2880	gi 642655		unknown [Rhizobium meliloti]	66	29	330
175	11	8743	7994	gi 854655		Na/H antiporter system [Bacillus alcalophilus]	66	43	730
190	5	7079	5727	gi 451072		di-tripeptide transporter [Lactococcus lactis]	66	40	1353
195	15	13919	13713	gi 1322411		unknown [Mycobacterium tuberculosis]	66	42	207
217	3	2822	2595	gi 1143542		alternative stop codon [Nectus norvegicus]	66	36	228
233	9	7133	6135	gi 1458327		F08F3.4 gene product [Caenorhabditis elegans]	66	47	999
218	1	41	1041	gi 809541		ChrA protein [Erwinia chrysanthemi]	66	42	994
241	1	2102	1053	gi 153067		peptidoglycan hydrolase [Staphylococcus aureus]	66	53	1050
261	1	1178	648	gi 1510859		M. jennaschii predicted coding region NJ0790 [Methanococcus jennaschii]	66	40	531
263	3	3731	2973	gi 1205865		tetrahydrodipicolinate N-succinyltransferase [Haemophilus influenzae]	66	47	759
272	8	6548	5484	gi 882101		high affinity nickel transporter [Alcaligenes eutrophus] ap P23516 HOXA_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.	66	44	1065
276	3	2805	2104	gi 1208965		hypothetical 23.3 kd protein [Escherichia coli]	66	47	702
278	2	2830	1784	gi 1488662		phosphatase-associated protein [Bacillus subtilis]	66	48	1047
278	3	3830	2952	gi 303560		ORF271 [Escherichia coli]	66	45	879
279	2	3894	2218	gi 1185289		2-succinyl-6-hydroxy-2,4-cyclohexadiene-1- carboxylate synthase [Bacillus subtilis]	66	48	1677
288	4	2535	2275	gi 1256625		putative [Bacillus subtilis]	66	42	261
292	2	1133	942	gi 1511604		M. jennaschii predicted coding region NJ1651 [Methanococcus jennaschii]	66	30	192

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
294	1	1116	gi 216314	esterase [Bacillus stearothermophilus]	66	45	558
297	4	2913	gi 994794	cytochrome a assembly facto [Bacillus subtilis] sp P24009 COXX_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	45	936
316	4	2053	gi 1107839	alginate lyase [Pseudomonas aeruginosa]	66	40	630
338	4	2460	gi 520750	biotin synthetase [Bacillus sphaericus]	66	59	159
339	1	1214	gi 467468	7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Bacillus subtilis]	66	52	480
363	1	863	gi 581649	epiC gene product [Staphylococcus epidermidis]	66	47	861
366	2	232	gi 1103505	unknown [Schizosaccharomyces pombe]	66	53	252
367	4	2468	sp P20692 TYRA_	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH)	66	50	624
372	3	2150	gi 467416	unknown [Bacillus subtilis]	66	38	552
378	1	212	gi 147309	purine nucleoside phosphorylase [Escherichia coli]	66	50	798
401	1	482	gi 388263	p-aminobenzoic acid synthase [Streptomyces griseus] pir JN0531 JN0531 p-aminobenzoic acid synthase - Streptomyces rissus	66	46	482
404	7	4826	gi 606784	cytidine deaminase [Bacillus subtilis]	66	51	429
411	2	1738	gi 1460081	unknown [Mycobacterium tuberculosis]	66	46	636
420	1	2	gi 1046024	[Na+ ATPase subunit J [Mycoplasma genitalium]	66	49	540
431	1	858	gi 1500008	M. jannaschii predicted coding region M3154 [Methanococcus jannaschii]	66	50	858
443	7	5679	gi 852076	MrgA [Bacillus subtilis]	66	46	381
444	3	3405	gi 153047	lysostaphin (ttg start codon) [Staphylococcus simulans] pir A25881 A25881 lysostaphin precursor - Staphylococcus simulans sp P10547 LSTP-STASI LYSOSTAPHIN PRECURSOR (EC 3.5.1.-)	66	51	993
561	1	956	gi 1204905	DNA-3-methyladenine glycosylase I [Haemophilus influenzae]	66	45	477
562	3	1066	gi 1046082	M. genitalium predicted coding region M0372 [Mycoplasma genitalium]	66	52	318
576	1	11	gi 105014	ORF_0234 [Escherichia coli]	66	43	714
577	3	1190	gi 1001353	hypothetical protein [Synecocystis sp.]	66	52	288
584	1	2	sp P24204 YEBA_	HYPOTHETICAL 46.7 KD PROTEIN IN MSBB-RUVB INTERGENIC REGION (ORF)	66	48	330
592	1	1410	gi 1928839	ORF266; putative [Lactococcus lactis phage BK5-T]	66	51	705
601	1	1433	gi 1488695	novel antigen; orf-2 [Staphylococcus aureus]	66	55	714

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
619	3	468	845	gi 746573	similar to M. musculus transport system membrane protein, Ncomp PIR:A40739 and S. cerevisiae SHF1 protein (PIR:A45154) Caenorhabditis elegans	66	45	378
706	2	561	355	gi 804808	unknown protein (Rattus norvegicus)	66	46	207
734	2	673	512	gi 1519085	phosphatidylcholine binding immunoglobulin heavy chain Igh variable region (Mus musculus)	66	60	162
740	1	3	317	gi 1209272	argininosuccinate-lyase (Campylobacter jejuni)	66	47	315
764	1	310	747	gi 435296	alkaline phosphatase like protein (Lactococcus lactis) pir S39339 S39339 alkaline phosphatase-like protein - Lactococcus lactis	66	42	438
852	1	338	171	gi 536955	CC Site No. 361 (Escherichia coli)	66	43	168
886	1	3	158	gi 289272	ferrichrome-binding protein (Bacillus subtilis)	66	44	156
889	1	462	232	gi 833061	HCMVUL77 (AA 1-642) Human cytomegalovirus	66	66	231
893	1	2	247	gi 149008	putative (Helicobacter pylori)	66	45	246
900	1	1425	733	gi 580842	F3 (Bacillus subtilis)	66	51	693
906	2	2300	1473	gi 790945	aryl-alcohol dehydrogenase (Bacillus subtilis)	66	53	828
947	1	79	549	gi 410117	diaminopimelate decarboxylase (Bacillus subtilis)	66	47	471
950	1	1100	552	gi 48713	orf145 (Staphylococcus aureus)	66	35	549
955	2	89	475	gi 1204390	uridine kinase (uridine monophosphokinase) (Haemophilus influenzae)	66	50	387
981	2	1308	997	gi 457146	rhostry protein (Plasmodium yoelii)	66	38	312
986	1	25	315	gi 105002	ORF_4356 (Escherichia coli)	66	31	291
1057	1	3	203	gi 1303853	YqgP (Bacillus subtilis)	66	40	201
1087	1	1	294	gi 575913	unknown (Saccharomyces cerevisiae)	66	53	294
1105	1	1	231	gi 1045799	methylgalactoside permease ATP-binding protein (Mycoplasma genitalium)	66	46	231
1128	1	2	574	gi 1001493	hypothetical protein (Synchocystis sp.)	66	46	573
1150	1	498	250	gi 1499034	M. jannaschii predicted coding region M0255 (Methanococcus jannaschii)	66	40	249
1180	2	707	453	gi 215908	DNA polymerase (g43) (Bacteriophage T4)	66	46	255
1208	1	1123	587	gi 1256653	DNA-binding protein (Bacillus subtilis)	66	58	537
1342	1	1	402	gi 1208474	hypothetical protein (Synchocystis sp.)	66	53	402
1761	2	589	398	gi 215811	tail fiber protein (Bacteriophage T3)	66	50	192

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
198J	1	499	251	gi11045935	DNA helicase II [Mycoplasma genitalium]	66	40	249
2103	2	176	400	gi129298	precursor for the major mezoelke surface antigens [Plasmodium alciaparum]	66	46	225
2341	1	373	188	gi1256623	exodeoxyribonuclease [Bacillus subtilis]	66	38	186
2458	1	325	184	gi11019410	unknown [Schizosaccharomyces pombe]	66	47	162
2505	1	468	235	gi11510394	putative transcriptional regulator [Methanococcus jannaschii]	66	39	234
2525	1	558	280	gi11000695	cytotoxin L [Clostridium ordellii]	66	44	279
2815	1	3	275	gi1765073	autolysin [Staphylococcus aureus]	66	47	273
3005	1	114	305	gi1205784	heterocyst maturation protein [Haemophilus influenzae]	66	46	192
3088	1	80	277	gi11303813	YgeW [Bacillus subtilis]	66	42	198
3071	1	1	189	gi11070014	protein-dependent [Bacillus subtilis]	66	41	189
3081	1	404	225	gi1984212	unknown [Schizosaccharomyces pombe]	66	44	180
3090	2	580	386	gi1204987	DNA polymerase III, alpha chain [Haemophilus influenzae]	66	48	195
3118	1	1	387	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	66	49	387
3739	1	798	400	gi11109684	ProV [Bacillus subtilis]	66	47	399
3796	1	402	202	gi1853760	acyl-CoA dehydrogenase [Bacillus subtilis]	66	60	201
3924	1	493	347	gi1563952	gluconate permease [Bacillus licheniformis]	66	46	249
4240	1	3	350	gi1151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir[A44756/A44756] hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	66	51	348
4604	1	7	234	pir[A26713]BMMC	hemocyanin subunit II - Atlantic horseshoe crab	66	46	228
4	9	8845	9750	gi1145646	cytB [Escherichia coli]	65	35	906
6	5	2708	3565	gi1887824	ORF_0310 [Escherichia coli]	65	47	858
11	1	1993	998	gi1143402	recombination protein (ttg start codon) [Bacillus subtilis] gi11303923 RecN [Bacillus subtilis]	65	44	996
15	7	2493	3524	gi11403126	cicD gene product [Alcaligenes eutrophus]	65	38	1032
18	3	1908	1372	gi1349187	acylttransferase [Saccharomyces cerevisiae]	65	50	537
21	3	1467	2492	gi1149518	phosphoribosyl anthranilate transferase [Lactococcus lactis] pir[S3126/S3126 anthranilate phosphoribosyltransferase (EC 4.2.1.18) - Lactococcus lactis subsp. lactis	65	52	1026
25	4	3374	4312	gi11502420	malonyl-CoA:Acyl carrier protein transacylase [Bacillus subtilis]	65	44	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
27	2	390	626	gi 1212729	YqjJ [Bacillus subtilis]	65	45	237
31	12	11040	10387	gi 509245	D-hydroxyisopropate dehydrogenase [Lactobacillus delbrueckii]	65	41	654
38	24	19172	19528	gi 547	H-protein [Flaveria croustetii]	65	41	357
44	2	790	1746	gi 405882	YekK [Escherichia coli]	65	46	957
44	12	9356	8832	gi 1205905	molybdenum cofactor biosynthesis protein [Haemophilus influenzae]	65	50	525
45	8	6635	7588	gi 493074	ApbA protein [Salmonella typhimurium]	65	46	954
51	2	580	1503	gi 580897	OppB gene product [Bacillus subtilis]	65	45	924
52	1	225	953	gi 1205518	NAD(P)H-flavin oxidoreductase [Haemophilus influenzae]	65	45	729
55	4	1339	1058	gi 444591A444	troponin T beta Tm-5 - rabbit	65	41	282
67	9	7421	8272	gi 143607	sporulation protein [Bacillus subtilis]	65	42	852
73	5	4466	5375	gi 1204896	lysophospholipase L2 [Haemophilus influenzae]	65	37	930
74	3	954	478	gi 1204844	H. influenzae predicted coding region H10594 [Haemophilus influenzae]	65	50	477
77	1	2	757	gi 1046082	M. genitalium predicted coding region M0372 [Mycoplasma genitalium]	65	46	756
77	2	795	1433	gi 1222116	permease [Haemophilus influenzae]	65	37	639
81	3	4728	3454	gi 1001708	hypothetical protein [Synecocystis sp.]	65	49	1275
91	7	8548	8357	gi 1399263	cystathionine beta-lyase [Emerella nidulans]	65	40	192
98	3	1608	1988	gi 467423	unknown [Bacillus subtilis]	65	38	381
98	4	2250	2987	gi 467424	unknown [Bacillus subtilis]	65	45	738
102	3	2598	2119	gi 1511532	N-terminal acetyltransferase complex, subunit ARD1 [Methanococcus jannaschii]	65	39	480
102	4	3647	2862	gi 1204637	H. influenzae predicted coding region H10388 [Haemophilus influenzae]	65	32	786
103	9	10851	5841	gi 142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	65	47	1011
103	10	10439	10119	gi 710021	nitrite reductase (nirD) [Bacillus subtilis]	65	51	321
106	2	262	1140	gi 39881	ORP 311 (AA 1-311) [Bacillus subtilis]	65	44	879
109	5	3909	4268	gi 1204399	glucosamine-6-phosphate deaminase protein [Haemophilus influenzae]	65	44	360
109	10	7165	8595	gi 536955	CG Site No. 361 [Escherichia coli]	65	41	1431

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
110	4	3688	3915	gi 407881	stringent response-like protein [Streptococcus equisimilis] p1r[S39975]S39975 stringent response-like protein - Streptococcus equisimilis	65	45	228
110	5	3882	4295	gi 407880	ORF1 [Streptococcus equisimilis]	65	50	414
110	6	4231	4380	gi 1139574	orf2 [Streptomyces griseus]	65	56	150
112	10	9218	8640	gi 1204571	H1 influenzae predicted coding region H10318 [Haemophilus influenzae]	65	52	579
112	12	12049	11288	gi 710496	transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	gi 1151158	repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	gi 37589	precursor [Homo sapiens]	65	46	420
127	11	10733	12658	gi 1064809	homologous to sp:HTRA_ECOL1 [Bacillus subtilis]	65	41	1926
143	8	7543	7004	gi 216513	mutator mutT [AT-GC transversion] [Escherichia coli]	65	56	540
145	5	3587	3838	gi 1209768	D02_orf569 [Mycoplasma pneumoniae]	65	37	642
150	4	3482	2841	gi 1146225	putative [Bacillus subtilis]	65	50	1911
166	1	3858	1948	gi 148304	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] p1r[A42296]A42296 lysostyme 2 [EC 3.2.1.-] precursor - Enterococcus irae (ATCC 9790)	65	27	252
188	6	3195	4178	gi 151943	ORF3; putative [Rhodobacter capsulatus]	65	46	984
189	9	4982	4785	gi 58812	ORF IV (NA 1-489) [Fluorimosaic virus]	65	40	198
195	6	7908	5272	gi 145220	alanyl-tRNA synthetase [Escherichia coli]	65	44	2617
195	7	10599	8104	gi 882711	exonuclease V alpha-subunit [Escherichia coli]	65	38	2496
206	16	16896	18191	gi 1468115	ornithine acetyltransferase [Bacillus subtilis]	65	53	1296
217	4	3844	3215	gi 1205974	5'-guanylate kinase [Haemophilus influenzae]	65	41	630
220	4	5265	3751	gi 580920	rodD [gtaA] polypeptide (AA 1-673) [Bacillus subtilis] p1r[S06048]S06048 probable rodD protein - Bacillus subtilis sp P13484 TACR_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHI-GLUCOSYLTRANSFERASE [EC 2.4.1.52] [TECHNOIC ACID BIOSYNTHESIS ROTININ E]	65	40	1515
236	5	2327	3709	gi 1146200	DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]	65	46	1383
237	3	1902	2513	gi 149379	HlsB [Lactococcus lactis]	65	46	612
241	4	4968	4195	gi 1205308	ribonuclease H1 [EC 3.1.264] [RNASE H1] [Haemophilus influenzae]	65	50	774
252	1	1278	940	gi 1206989	hypothetical protein [CB-U00022.9] [Haemophilus influenzae]	65	40	339
261	5	4780	3794	gi 145927	fecD [Escherichia coli]	65	43	987

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
274	1	3	278	gi 496558	orfX [Bacillus subtilis]	65	42	276
301	2	982	815	gi 167418	unknown [Bacillus subtilis]	65	45	168
307	4	1586	2884	gi 1070014	protein-dependent [Bacillus subtilis]	65	40	723
335	2	2286	1399	gi 146913	N-acetylglucosamine transport protein [Escherichia coli] pir[B29895]MOEC2N phosphotransferase system enzyme II (EC 7.1.69), N-acetylglucosamine-specific - Escherichia coli sp P09333 PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC COMPONENT (EIIA)	65	50	888
338	5	4120	3170	gi 1277029	biotin synthase [Bacillus subtilis]	65	49	951
343	3	1490	2800	gi 143264	membrane-associated protein [Bacillus subtilis]	65	48	1311
344	4	2761	2531	gi 1050540	tRNA-glutamine synthetase [Lupinus luteus]	65	34	231
358	3	3421	3621	gi 1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	65	47	201
364	1	238	699	gi 1340128	ORF1 [Staphylococcus aureus]	65	51	462
379	1	1	576	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir A27650 A27650 regulatory protein phor - Bacillus subtilis sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-)	65	40	576
379	3	3666	4346	gi 143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) [Bacillus subtilis]	65	50	681
428	1	187	483	gi 1420465	ORF YOR195w [Saccharomyces cerevisiae]	65	45	297
438	2	272	838	gi 143498	dugS protein [Bacillus subtilis]	65	18	567
444	11	9280	10215	gi 1204756	ribokinase [Haemophilus influenzae]	65	47	936
449	2	1241	1531	gi 598848	Na/H antiporter homolog [Lactococcus lactis]	65	41	291
478	2	1452	865	gi 1045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	65	39	588
479	1	1032	517	gi 1498192	putative [Pseudomonas aeruginosa]	65	40	516
480	6	4312	5637	gi 415662	UDP-N-acetylglucosamine 1-carboxyvinyl transferase [Acinetobacter alcoaceticus]	65	48	1326
484	1	2	430	gi 146551	transmembrane protein (kdpD) [Escherichia coli]	65	44	429
499	1	54	932	gi 603456	reductase [Leishmania major]	65	53	879
505	1	914	459	gi 1518853	OafA [Salmonella typhimurium]	65	39	456
571	2	1509	883	gi 49399	open reading frame upstream glnE [Escherichia coli] ir[S37754]S37754 hypothetical protein XE (glnE 5' region) - Escherichia coli	65	44	627
611	2	506	270	gi 10961	RAP-2 [Plasmodium falciparum]	65	40	237

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
705	1	564	283	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	65	52	282
712	1	1	177	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	177
712	2	196	354	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	159
743	1	2	631	gi 310631	ATP binding protein (Streptococcus gordonii)	65	45	630
749	2	393	779	gi 467374	single strand DNA-binding protein [Bacillus subtilis]	65	29	387
762	1	1698	850	gi 160399	multidrug resistance protein [Plasmodium falciparum]	65	48	849
788	1	85	315	gi 1129096	unknown protein [Bacillus sp.]	65	35	231
850	1	1	408	gi 1006604	hypothetical protein [Synechocystis sp.]	65	37	408
908	1	1	444	gi 1199546	[2362] Saccharomyces cerevisiae	65	46	444
925	1	1	174	gi 1256653	DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	232	gi 238657	AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, eptide, 514 aa]	65	47	207
1037	1	414	262	gi 1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1053	1	348	175	gi 642655	unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	gi 1162980	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	65	48	648
1214	1	881	495	gi 1205959	lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	pir S35493 S354	site-specific DNA-methyltransferase Stet (EC 2.1.1.-) - Streptococcus sanguis	65	35	201
1276	2	900	577	gi 473794	ORF' [Escherichia coli]	65	34	324
2057	1	272	138	gi 633699	TrsH [Yersinia enterocolitica]	65	21	135
2521	1	336	169	gi 1045789	hypothetical protein (CB:U14003_76) [Mycoplasma genitalium]	65	41	168
2974	1	590	297	gi 152052	enantiomerase-selective amidase [Rhodococcus sp.]	65	45	294
3031	1	306	154	pir JQ1024 JQ10	hypothetical 30K protein (Dmnp140 5' region) - fruit fly (Drosophila melanogaster)	65	45	153
3069	1	3	278	gi 144906	product homologous to E.coli thioredoxin reductase: J.Biol.Chem. 1988; 263:9015-9019, and to F52a protein of alkyl hydroperoxide oxidase from S.typhimurium: J.Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	65	46	276
3146	1	282	142	gi 49315	ORF1 gene product [Bacillus subtilis]	65	47	141
3170	1	679	341	gi 1507711	indolepyruvate decarboxylase [Erwinia herbicola]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3546	1	1	303	gi 450688	hsdM gene of EcoRII gene product [Escherichia coli] pIC[S38437] hsdM protein - Escherichia coli pir[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-570)	65	42	303
3782	1	2	328	gi 166412	[NADH-glucamate synthase (Medicago sativa)]	65	42	327
3990	1	374	189	gi 1009366	[Respiratory nitrate reductase (Bacillus subtilis)]	65	53	186
4032	1	613	308	gi 1323127	[ORF YGR087c (Saccharomyces cerevisiae)]	65	50	306
4278	2	726	364	gi 1197667	[vitellogenin (Anolis pulchellus)]	65	42	363
19	4	4259	5518	gi 145727	[dead (Escherichia coli)]	64	45	1260
19	6	7639	6926	gi 1016232	[ycf27 gene product (Cyanophora paradoxa)]	64	36	714
20	8	7053	6454	gi 765073	[autolysin (Staphylococcus aureus)]	64	47	600
31	13	12706	11537	gi 414009	[ipa-85d gene product (Bacillus subtilis)]	64	45	1170
33	4	2388	4364	gi 1204696	[fructose-permease IFBC component (Haemophilus influenzae)]	64	47	1977
36	3	1871	3013	gi 290503	[glutamate permease (Escherichia coli)]	64	40	1143
37	6	4065	4409	gi 39815	[orf 2 gene product (Bacillus subtilis)]	64	46	345
45	9	7852	8760	gi 1230585	[nucleotide sugar epimerase (Vibrio cholerae O139)]	64	53	909
53	3	1540	1899	gi 1303961	[yqjJ (Bacillus subtilis)]	64	50	360
56	6	4793	3855	gi 457514	[glcC (Bacillus subtilis)]	64	45	919
56	24	30002	30247	gi 470331	[similar to zinc fingers (Caenorhabditis elegans)]	64	42	246
62	4	2759	2421	gi 642655	[unknown (Rhizobium meliloti)]	64	28	339
85	6	7178	6027	gi 457702	[5-aminimidazole ribonucleotide-carboxylase (Pichia methanolica) pIC[S39112]S39112 phosphoribosylaminimidazole carboxylase (EC 1.1.21) - yeast (Pichia methanolica)]	64	46	1152
96	9	9251	10030	gi 1511513	[ABC transporter, probable ATP-binding subunit (Methanococcus jannaschii)]	64	42	780
100	1	1	600	gi 765073	[autolysin (Staphylococcus aureus)]	64	44	600
106	5	3868	4854	gi 466778	[lysine specific permease (Escherichia coli)]	64	46	987
123	2	838	554	gi 467484	[unknown (Bacillus subtilis)]	64	47	285
127	8	7514	7810	gi 210061	[serotype-specific antigen (African horse sickness virus) pIC[S27891]S27891 capsid protein VP2 - African horse sickness virus]	64	28	297
131	7	7134	6721	gi 1511160	[M. jannaschii predicted coding region MJ1163 (Methanococcus jannaschii)]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
142	5	5455	4817	gi 1173517	riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	639
143	1	709	356	ptr A32950 A329	probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi 398151	major surface antigen MSG2 [Pneumocystis carinii]	64	44	261
154	4	3134	2307	gi 984587	dlnP [Escherichia coli]	64	50	828
161	5	3855	4880	gi 903304	ORF72 [Bacillus subtilis]	64	37	1026
165	1	33	791	gi 467483	unknown [Bacillus subtilis]	64	38	759
175	6	6355	4714	gi 1072398	phad gene product [Rhizobium meliloti]	64	42	1512
188	3	2042	2500	gi 1001961	MHC class II analog [Staphylococcus aureus]	64	45	459
195	14	13667	13446	gi 396380	No definition line found [Escherichia coli]	64	47	222
206	15	16429	16938	gi 304134	larcC [Bacillus stearothermophilus]	64	49	510
215	1	560	282	gi 142359	ORF 6 [Acetobacter vinelandii]	64	39	279
243	7	7818	6928	gi 414014	lpe-90d gene product [Bacillus subtilis]	64	49	891
258	2	1330	845	gi 664754	p17 [Listeria monocytogenes]	64	38	486
259	1	462	232	gi 1499663	M. jannaschii predicted coding region NJ0837 [Methanococcus jannaschii]	64	52	231
263	6	6565	5567	gi 142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp Q04797 DHAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC 2.1.1.11) [ASA DEHYDROGENASE]	64	48	999
271	1	3	1163	gi 467091	hflX; B2235_C2_202 [Mycobacterium leprae]	64	44	1161
280	1	173	1450	gi 1303839	YqfR [Bacillus subtilis]	64	43	1278
293	1	2532	1267	gi 147345	primosomal protein n' [Escherichia coli]	64	45	1266
295	2	742	1488	gi 459266	Potential membrane spanning protein [Staphylococcus hominis] pir S42932 S42932 potential membrane spanning protein - taphylococcus hominis	64	39	747
301	5	1625	1446	gi 580835	lysine decarboxylase [Bacillus subtilis]	64	35	180
315	4	5064	3949	gi 143396	guinol oxidase [Bacillus subtilis]	64	45	1116
321	1	1264	635	gi 710496	transcriptional activator protein [Bacillus brevis]	64	41	630
333	5	4520	4239	gi 1314295	ORF2; putative 19 kDa protein [Listeria monocytogenes]	64	43	282
342	1	1	549	gi 142940	ftsA [Bacillus subtilis]	64	38	549
353	3	2878	2324	gi 537049	ORF_o470 [Escherichia coli]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	2	827	3458	pir S25295 A328	oxoglutarate dehydrogenase (lipomide) (EC 1.2.4.2) - Bacillus subtilis	64	47	2332
404	6	4429	4839	pir A36933 A369	diacylglycerol kinase homolog - Streptococcus mutans	64	35	411
407	1	2020	1133	gi 969026	OrfX [Bacillus subtilis]	64	41	888
425	1	1109	591	gi 1146177	phosphotransferase system glucose-specific enzyme II (Bacillus subtilis)	64	44	519
443	6	4082	4798	gi 147309	purine nucleoside phosphorylase [Escherichia coli]	64	51	717
450	2	1035	1604	gi c06376	ORP_0162 [Escherichia coli]	64	38	570
470	5	1680	6107	gi 1369948	host interacting protein [Bacteriophage B1]	64	45	4428
486	4	1911	1471	gi 1205582	spermidine/putrescine transport system permease protein [Haemophilus influenzae]	64	35	441
497	1	2217	1159	sp P36929 FHU_E	FHU PROTEIN	64	38	1059
501	1	3	410	gi 142450	shrC protein [Bacillus subtilis]	64	38	408
514	1	3	290	gi 1204496	H. influenzae predicted coding region H10238 [Haemophilus influenzae]	64	34	288
551	4	3162	3323	gi 1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	64	41	162
603	4	759	956	gi 755823	NADH dehydrogenase F [Streptomyces americana]	64	35	198
653	2	940	746	gi 1213234	dicarboxylic amino acids Dip5 permease [Saccharomyces cerevisiae]	64	41	195
660	3	3401	2257	sp P46133 YDAUL	HYPOTHETICAL PROTEIN IN OGT 5' REGION (FRAGMENT)	64	39	1345
695	1	11	502	gi 1001383	hypothetical protein [Synecocystis sp.]	64	41	492
702	1	3	752	gi 142865	DNA primase [Bacillus subtilis]	64	46	750
826	1	1	339	gi 971336	arginyl tRNA synthetase [Bacillus subtilis]	64	50	339
838	1	1831	917	gi 1354775	pfos/R [Treponema pallidum]	64	41	915
864	3	675	944	gi 39833	cyclomaltodextrin glucanotransferase [Bacillus stearothermophilus] i 39835	64	47	270
887	1	3	677	gi 153002	enterotoxin type E precursor [Staphylococcus aureus] pir A28179 A28179 enterotoxin E precursor - Staphylococcus aureus sp P12993 ETXE_STAAU ENTEROTOXIN TYPE E PRECURSOR (SEE)	64	46	675
928	2	1172	963	gi 311976	fibrinogen-binding protein [Staphylococcus aureus] pir S34270 S34270 fibrinogen-binding protein - Staphylococcus aureus	64	41	210
1049	2	800	606	gi 1049115	Pap60 [Bacillus subtilis]	64	42	195
1067	2	999	748	gi 1151072	hhdA precursor [Haemophilus ducreyi]	64	50	252

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1120	1	50	202	gi1142439	ATP-dependent nuclease [Bacillus subtilis]	64	30	153
1125	1	751	377	gi1581648	epiB gene product [Staphylococcus epidermidis]	64	44	375
1688	1	402	216	pir1A013651TVNS	transforming protein K-raa - mouse	64	47	189
2472	1	2	358	gi1487282	Na ⁺ -ATPase subunit J [Enterococcus hirae]	64	36	357
2989	1	520	356	gi1304134	argC [Bacillus stearothermophilus]	64	50	165
3013	1	630	352	gi1551699	cytochrome oxidase subunit I [Bacillus firmus]	64	51	279
3034	1	546	274	gi1204349	hypothetical protein (GB:GBD0212.3) [Haemophilus influenzae]	64	50	273
3197	1	613	308	gi11009366	Respiratory nitrate reductase [Bacillus subtilis]	64	46	306
3303	1	90	362	gi1107839	alginate lyase [Pseudomonas aeruginosa]	64	43	273
3852	2	82	288	gi1216746	D-lactate dehydrogenase [Lactobacillus plantarum]	64	42	207
3868	1	1	312	gi1149435	putative [Lactococcus lactis]	64	48	312
3918	1	660	331	gi15532	acetyl-CoA acyltransferase [Yarrowia lipolytica]	64	46	330
4000	1	112	378	gi1944688	unknown [Saccharomyces cerevisiae]	64	44	267
4009	1	81	368	gi139372	grsB gene product [Bacillus brevis]	64	41	288
4166	1	2	349	gi1149435	putative [Lactococcus lactis]	64	46	348
4366	1	2	307	gi1216267	ORF2 [Bacillus megaterium]	64	44	306
4457	1	2	400	gi1197667	vitellogenin [Anolis pulchellus]	64	43	399
11	3	1539	2438	gi1438228	ORF C [Staphylococcus aureus]	63	32	900
24	7	5611	5423	gi11369943	a1 gene product [Bacteriophage B1]	63	34	189
29	1	1	390	gi1467441	expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi1467441 expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed Bacil	63	43	390
31	6	6329	5712	gi1496943	ORF [Saccharomyces cerevisiae]	63	47	618
44	23	14669	15019	pir1A044461OOC	hypothetical protein P-92 - Escherichia coli	63	36	351
48	6	4403	6250	gi143498	pyruvate synthase [Halobacterium halobium]	63	42	1848
50	5	3869	4738	gi1413967	ipa-43d gene product [Bacillus subtilis]	63	43	870
53	6	6764	5742	gi1474176	regulator protein [Staphylococcus xylosus]	63	49	1023

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match	Accession	Match gene name	% sim	% ident	Length (nt)
56	14	15880	17607	gi 467409		DNA polymerase III subunit [Bacillus subtilis]	63	46	1728
57	11	7945	7376	gi 137036		ORF_0158 [Escherichia coli]	63	39	570
62	3	2479	2114	gi 42656		unknown [Rhizobium meliloti]	63	41	366
70	8	6562	7353	gi 1399821		PhoC [Rhizobium meliloti]	63	46	792
75	2	223	927	gi 149376		HISG [Lactococcus lactis]	63	45	705
78	5	4912	4403	gi 413950		lpa-26d gene product [Bacillus subtilis]	63	42	510
91	5	9076	7220	gi 466997		metH2; B2126-CL157 [Mycobacterium leprae]	63	41	1857
91	8	10566	9448	gi 1204344		cystathionine gamma-synthase [Haemophilus influenzae]	63	45	1119
120	1	21	1508	gi 882657		sulfite reductase (NADPH) flavoprotein beta subunit [Escherichia coli]	63	46	1488
120	4	2722	4125	gi 665994		hypothetical protein [Bacillus subtilis]	63	34	1404
127	7	6064	7566	gi 40162		murE gene product [Bacillus subtilis]	63	44	1503
149	6	2321	2106	gi 148503		dnaK [Erysipelothrix rhusiopathiae]	63	40	216
149	26	10445	10170	gi 4870		ORF 2, has similarity to DNA polymerase (Saccharomyces kluyveri) r[S15961][S15961 hypothetical protein 2 - yeast (Saccharomyces yverii) plasmid pSKL	63	42	276
164	2	507	1298	gi 145476		CDP-diglyceride synthetase [Escherichia coli]	63	44	792
166	6	9909	8164	gi 151932		fructose 6-phosphate 1-phosphotransferase [Rhodospirillum rubrum]	61	41	1746
169	4	1704	1886	gi 152886		elongation factor Ts (tsf) [Spiroplasma citri]	63	48	183
188	5	3145	2951	gi 1334547		GIV COI 114 grp IB protein [Podospora anserina]	63	42	195
195	13	11767	12804	gi 606100		ORF_0335 [Escherichia coli]	63	40	1038
201	2	607	2283	gi 433534		arginyl-tRNA synthetase [Corynebacterium glutamicum] p1r[A09936][A49936] arginine-tRNA ligase (EC 6.1.1.19) - oryzae bacterium glutamicum	63	46	1677
206	14	15893	16489	gi 580828		N-acetyl-glutamate-gamma-semialdehyde dehydrogenase [Bacillus subtilis]	63	49	597
220	5	7769	5766	gi 216334		secA protein [Bacillus subtilis]	63	42	2004
221	1	74	907	gi 677945		AppA [Bacillus subtilis]	63	42	834
227	3	944	1708	gi 1510558		isobutyrate synthase [Methanococcus jannaschii]	63	46	765
261	2	804	1070	gi 486511		ORF YKR054c [Saccharomyces cerevisiae]	63	45	267
269	2	3606	1960	gi 148221		DNA-dependent ATPase, DNA helicase [Escherichia coli] p1r[JS0137][BVECRQ] recQ protein - Escherichia coli	63	42	1647

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
278	8	7417	6176	gi 699273	cystathionine gamma-synthase [Mycobacterium leprae] sp P46807 MEFB_MYCLE_CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9) O-SUCCINYLMONOSERINE (THIOLE)-LYASE]	63	41	1242
287	2	738	1733	gi 405133	putative [Bacillus subtilis]	63	38	996
295	1	2	748	gi 1239983	hypothetical protein [Bacillus subtilis]	63	41	747
328	3	2148	3134	gi 45302	carrier protein (AA 1 - 437) [Pseudomonas aeruginosa] ir S11497 S11497 branched-chain amino acid tRNA:port protein braB - eudomonas aeruginosa	63	36	987
362	2	1026	1216	sp P5136 SERA_	D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PCDH)	63	38	411
404	1	326	1051	gi 1303816	Yqe2 [Bacillus subtilis]	63	35	726
405	3	2101	1715	gi 1303914	YqjV [Bacillus subtilis]	63	42	387
406	1	451	227	gi 1412152	sulfate permease (99 start codon) [Synecococcus PCC6301] pif A30301 CRYS7 sulfate transport protein - Synecococcus sp. PCC 7942)	63	43	225
415	2	1048	2718	gi 1205402	transport ATP-binding protein [Haemophilus influenzae]	63	41	1671
426	4	3575	2679	gi 393268	29-kilodalton protein [Streptococcus pneumoniae] sp P42362 P29K_STRPN 29 KD MEMBRANE PROTEIN IN PSA 5'REGION ORF1)	63	39	897
505	3	1347	2195	gi 1418999	orf4 [Lactobacillus sake]	63	40	849
507	1	2	574	gi 546917	comK [Bacillus subtilis, E26, Peptide, 192 aa]	63	35	573
562	2	146	1084	gi 43985	nifs-like gene [Lactobacillus delbrueckii]	63	45	919
675	1	427	215	gi 1510994	serine aminotransferase [Methanococcus jannaschii]	63	29	213
686	1	3	230	gi 517356	nitrate reductase (NADH) [Lotus japonicus]	63	52	228
701	1	3	392	gi 881940	NorQ protein [Paracoccus denitrificans]	63	41	390
720	1	2	400	gi 47168	open reading frame [Streptomyces lividans]	63	35	399
779	1	571	287	gi 1261932	unknown [Mycobacterium tuberculosis]	63	41	285
907	1	22	321	gi 149445	ORF1 [Lactococcus lactis]	63	27	300
972	1	794	399	gi 1511235	M. jannaschii predicted coding region MJ1232 [Methanococcus jannaschii]	63	27	396
1085	1	1154	618	gi 1204277	hypothetical protein (CB:U00019_14) [Haemophilus influenzae]	63	38	537
1094	1	3	542	gi 790943	urea amidolyase [Bacillus subtilis]	63	39	540
1108	1	3	482	gi 549892 S498	regulation protein - Bacillus subtilis	63	44	480
1113	1	1231	617	gi 493017	endocarditis specific antigen [Enterococcus faecalis]	63	45	615

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	matc gene name	% sim	% ident	length (nt)
1300	1	3	695	[sp P31940 YQJH_	HYPOTHETICAL 54.3 KO PROTEIN IN ECO-ALKB INTERGENIC REGION.	63	46	693
1335	1	1	204	[gi 928989	p100 protein [Borrelia burgdorferi]	63	30	204
1814	1	3	245	[gi 1303914	[YQNY [Bacillus subtilis]	63	34	243
2021	1	498	250	[pir C33496 C334	[hisC homolog - Bacillus subtilis	63	46	249
2335	1	2	193	[gi 436132	[product is similar to TnpA of transposon Tn554 from Staphylococcus aureus [Clostridium butyricum]	63	40	192
2335	1	1	195	[gi 1184298	[flagellar MS-ring protein [Borrelia burgdorferi]	63	47	195
2406	1	451	227	[gi 1041785	[rhopty protein [Plasmodium yoelii]	63	33	225
2961	2	136	360	[gi 312443	[carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldylyticus]	63	52	325
2965	1	1	402	[gi 140784	[orf-1; novel antigen [Staphylococcus aureus]	63	50	402
2987	1	583	293	[gi 1224069	[amidase [Moraxella catarrhalis]	63	35	291
2994	1	266	135	[gi 836646	[phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeoidea]	63	51	132
3043	1	440	252	[gi 1480237	[phenylacetaldehyde dehydrogenase [Escherichia coli]	63	40	189
3078	1	609	400	[gi 1148792	[intrinsic membrane protein [Mycoplasma hominis]	63	36	210
3139	1	2	217	[gi 439126	[glutamate synthase (NADPH) [Acetivibrium brasilense] pir[A49916]A49916	63	47	216
3625	1	79	198	[gi 621073	[GMP100; putative [Bacteriophage M1-III]	63	48	196
3658	1	1	399	[gi 1301697	[YrKA [Bacillus subtilis]	63	37	399
3659	1	3	395	[gi 1256135	[YbbP [Bacillus subtilis]	63	48	393
3783	1	720	361	[gi 1256902	[Pyruvate decarboxylase isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae]	63	34	360
3900	1	338	171	[sp P10537 AHVB	[BETA-AMYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOTRIHYDROLASE) .	63	54	168
4309	1	3	176	[pir A37967 A379	[neural cell adhesion molecule Ng-CAM precursor - chicken	63	57	174
4367	1	1	195	[gi 1121932	[Par6p gene product [Pichia pastoris]	63	30	195
4432	1	1	312	[gi 1351259	[HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir[A44756]A44756	63	51	312
4488	1	6	308	[gi 296464	[ATPase [Lactococcus lactis]	63	36	303
33	3	1411	2400	[gi 1153675	[tagatose 6-P kinase [Streptococcus mutans]	62	44	990
36	9	5985	6218	[gi 1490521	[HMSH] [Homo sapiens]	62	51	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
37	1	2	721	gi 1107531	ceuE gene product (Campylobacter coli)	62	33	720
38	15	110912	111589	gi 1222058	H. influenzae predicted coding region HIN1279 (Haemophilus influenzae)	62	38	678
38	25	119526	20329	gi 695280	ORF2 (Alicapnia eutrophus)	62	41	804
57	2	2523	1780	gi 171234	orf1 (Haemophilus influenzae)	62	55	744
57	9	6646	6350	gi 508174	EF1B domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	62	35	297
58	1	2	559	gi 755152	highly hydrophobic integral membrane protein (Bacillus subtilis) spIP42953 TAGC_BACSU TEICHOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGG.	62	34	558
67	110	8250	9014	gi 470683	Shows similarity with ATP-binding proteins from other ABC-transport perons. Swiss Prot Accession Numbers P24137, P08007, P04285, P24136 Escherichia coli	62	34	765
69	8	8315	7494	gi 46816	actVA 4 gene product (Streptomyces coelicolor)	62	44	822
80	3	1793	1320	gi 39993	UDP-N-acetylmuramoylalanine--D-glutamate ligase (Bacillus subtilis)	62	43	474
87	7	7034	9205	gi 217191	5'-nucleotidase precursor (Vibrio parahaemolyticus)	62	48	2172
100	3	4051	3089	gi 1511047	phosphoglycerate dehydrogenase (Methanococcus jannaschii)	62	42	963
102	1	2	520	gi 153655	mismatch repair protein (Streptococcus pneumoniae) pir[C28667 C28667 DNA mismatch repair protein hexA - Streptococcus pneumoniae	62	34	519
112	2	466	1068	gi 153741	ATP-binding protein (Streptococcus mutans)	62	37	603
114	7	6855	7562	gi 1204866	L-fucose operon activator (Haemophilus influenzae)	62	38	708
116	4	6823	5633	gi 677947	AppC (Bacillus subtilis)	62	37	1191
124	8	6855	6004	gi 853777	product similar to E. coli PRA2 protein (Bacillus subtilis) pir 553438 553438 ynfE protein (Bacillus subtilis sp P45873 HEMK_BACSU POSSIBLE PROTOPORPHYRINOGEN OXIDASE (EC 1.3.1.1).	62	44	852
148	1	24	554	gi 467456	unknown (Bacillus subtilis)	62	30	531
149	20	7591	6725	gi 1203807	replicative DNA helicase (Haemophilus influenzae)	62	41	867
163	3	1503	1153	gi 40067	X gene product (Bacillus anthracis)	62	42	351
164	15	114673	15632	gi 42219	P15 gene product (AA 1 - 314) (Escherichia coli)	62	38	960
165	2	1166	1447	gi 403936	phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector	62	38	282
166	2	2084	5089	gi 308861	GTC start codon (Lactococcus lactis)	62	44	3006
171	1	1225	614	gi 1046053	hypothetical protein (SP:P32049) (Mycoplasma genitalium)	62	41	612

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi 143045	hemY [Bacillus subtilis]	62	45	1212
200	1	3	956	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	62	32	954
237	2	935	1966	gi 141695	hisc protein [Escherichia coli]	62	44	1032
261	3	4008	2605	gi 143121	ORF A; putative [Bacillus firmus]	62	42	1404
299	8	4477	4719	gi 467441	expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi 467441 expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed Bacil	62	47	243
304	6	5018	3819	gi 153015	FemA protein [Staphylococcus aureus]	62	43	1200
324	1	2	262	gi 142717	cytochrome a33 controlling protein [Bacillus subtilis] pir A33960 A33960 cta protein - Bacillus subtilis sp P12946 C7AA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN.	62	30	261
335	2	269	1207	gi 581088	methionyl-tRNA formyltransferase [Escherichia coli]	62	39	939
332	6	4894	4631	gi 1499960	uridine 5'-monophosphate synthase [Methanococcus jannaschii]	62	36	264
335	1	2	370	gi 145925	fecB [Escherichia coli]	62	32	369
365	8	6628	6804	gi 413943	lipo-19d gene product [Bacillus subtilis]	62	54	177
369	2	2744	1626	pir A43577 A435	regulatory protein pfor - Clostridium perfringens	62	42	1119
370	1	34	264	gi 40665	beta-glucosidase [Clostridium thermocellum]	62	37	331
415	3	2709	3176	gi 3205401	transport ATP-binding protein [Haemophilus influenzae]	62	35	468
429	1	1578	790	gi 1046024	Na+ ATPase subunit J [Myco plasma genitalium]	62	40	789
444	2	704	1369	gi 581510	modulation gene; integral membrane protein; homology to Rhizobium eguminosarum nod [Rhizobium lotii]	62	37	666
477	2	751	1869	pir A8440 A484	ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum	62	44	1119
485	1	241	1707	gi 117934	betaine aldehyde dehydrogenase [Beta vulgaris]	62	43	1467
487	3	1141	1311	gi 149445	ORF1 [Lactococcus lactis]	62	31	171
494	2	1134	1313	gi 166835	ribulose biphosphate carboxylase/oxygenase activase [Arabidopsis haliana]	62	37	180
518	1	193	882	gi 153491	O-methyltransferase [Streptomyces glaucosens]	62	39	690
534	2	369	2522	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	62	35	2154
551	6	4371	4820	gi 511113	ferric uptake regulation protein [Campylobacter jejuni]	62	37	450
574	1	1	570	gi 153000	enterotoxin B [Staphylococcus aureus]	62	43	570

TABLE 2

5 aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
590	2	344	1171	gi140367	ORF1 (Clostridium acetobutylicum)	62	37	828
655	1	396	830	gi1147195	pnb protein [Escherichia coli]	62	44	435
656	1	2	478	gi11205451	cell division inhibitor [Haemophilus influenzae]	62	36	477
676	1	692	348	gi11511613	methy coenzyme H reductase system, component A2 [Methanococcus jannaschii]	62	36	345
687	1	493	248	gi149272	Asparaginase [Bacillus licheniformis]	62	48	246
700	2	267	944	gi11205822	hypothetical protein (CB:X75627_4) [Haemophilus influenzae]	62	40	678
840	2	1715	1041	gi11045865	M. genitalium predicted coding region HG181 [Mycoplasma genitalium]	62	36	675
864	4	898	1491	gi11144332	deoxyuridine nucleotidohydrolase [Homo sapiens]	62	38	594
916	1	35	400	gi1413931	ipa-7d gene product [Bacillus subtilis]	62	45	366
1071	1	1	771	gi11510649	aspartokinase I [Methanococcus jannaschii]	62	40	771
1084	1	19	609	gi1688011	AQX-1 antigen [human, infertile patient, testis, Peptide, 505 aa]	62	39	591
1103	1	3	203	gi1581261	ORF homologous to E. coli metB [Herpetosiphon aurantiacus] pIR[S14030]S14030	62	51	201
1217	1	463	233	gi1460025	hypothetical protein - Herpetosiphon aurantiacus (fragment)	62	41	231
1533	1	644	414	gi1413968	ORF2, putative [Streptococcus pneumoniae]	62	48	231
1537	1	3	257	gi11510641	lne-44d gene product [Bacillus subtilis]	62	29	255
2287	1	3	161	gi1485956	alanyl-tRNA synthetase [Methanococcus jannaschii]	62	45	159
2386	1	3	245	gi1285708	lncC gene product [Proteus mirabilis]	62	31	243
2484	1	331	167	gi1142092	non-toxic component [Clostridium botulinum]	62	35	165
2490	1	798	400	gi1581648	DNA-repair protein (recA) [Anabaena variabilis]	62	42	399
3016	1	596	300	gi1710023	epiB gene product [Staphylococcus epidermidis]	62	51	297
3116	1	1	213	gi1466883	uroporphyrinogen III [Bacillus subtilis]	62	40	213
3297	1	823	413	gi1475715	nifs; BL496_C2_193 [Mycobacterium leprae]	62	42	411
3609	1	31	276	gi11408301	acetyl coenzyme A acetyltransferase (thiolase) [Clostridium acetobutylicum]	62	48	246
3665	2	584	402	gi1151259	homologous to N-acetyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	62	40	183
3733	1	3	374	gi11353197	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonil] pIR[A44756]A44756	62	42	372
					hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.			
					thioredoxin reductase [Eubacterium acidaminophilum]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3898	1	1	237	gi153675	tagatose 6-P kinase [Streptococcus mutans]	62	45	237
4027	1	283	143	gi1330705	homologue to gene 30 (aa 1-59); putative [Bovine herpesvirus 4]	62	43	141
4109	1	727	365	gi141748	hadW protein (AA 1-520) [Escherichia coli]	62	45	363
4303	1	1	303	gi1303813	Yqew [Bacillus subtilis]	62	43	303
4380	1	530	267	gi1215684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi1510692	enterotoxin H [Staphylococcus aureus]	62	34	255
4598	1	411	223	gi1763513	ORF4; putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi141748	hadW protein (AA 1-520) [Escherichia coli]	62	45	222
5	5	4288	3912	gi1928831	ORF95; putative [Lactococcus lactis phage BK5-7]	61	36	357
11	1	320	162	gi133356[C333	prothymosin alpha homolog (clone 32) - human (fragment)	61	33	159
16	11	10291	11938	gi11205391	hypothetical protein (SP:P3395) [Haemophilus influenzae]	61	44	948
32	1	281	801	gi11066504	exo-beta 1,3 glucanase [Cochliobolus carbonum]	61	50	519
38	3	616	1107	gi1510864	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	61	41	492
45	4	3082	4038	gi1109686	ProX [Bacillus subtilis]	61	45	957
48	8	7118	7504	gi1498839	ORF2 [Clostridium perfringens]	61	33	387
51	9	4605	5570	gi1388269	traC [Plasmid pAD1]	61	42	966
60	6	1689	2243	gi1205893	hypothetical protein (GB:U00011_3) [Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi1854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	61	38	438
67	5	4330	5646	gi1466612	nika [Escherichia coli]	61	36	1317
74	2	2400	1504	gi1204846	carbamate kinase [Haemophilus influenzae]	61	40	897
85	1	2198	1101	gi1498756	amidophosphoribosyltransferase PurF [Rhizobium etli]	61	41	1098
86	4	1995	1582	gi1499931	M. jannaschii predicted coding region M31083 [Methanococcus jannaschii]	61	44	414
97	1	74	649	gi1518679	orf [Bacillus subtilis]	61	44	576
99	2	2454	1990	gi1413958	ipa-34d gene product [Bacillus subtilis]	61	18	465
124	7	6223	5123	gi1556881	Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] gi1549358 S49358 ipc-29d protein - Bacillus subtilis ap p19153 YMLC_BACSU HYPOTHEICAL 37.0 KD PROTEIN IN SPO1R-GLYC NTERGENIC REGION.	61	46	1101
125	4	1668	2531	gi1491643	ORFA gene product [Chloroflexus aurantiacus]	61	43	864

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Accession	% sim	% ident	length (nt)
132	1	1250	627	pir[P00759]	hypothetical protein I - Enterococcus faecalis plasmid pAM-beta-1 (fragment)	61	43	624
149	9	3617	3075	gi1144132	deoxyuridine nucleotidohydrolase [Homo sapiens]	61	40	543
149	22	8690	7869	gi1160047	p101/acidic basic repeat antigen [Plasmodium falciparum] p1r[A29222][A29232] 101K malaria antigen precursor - Plasmodium falciparum (strain Camp)	61	35	822
168	3	1915	2361	gi11499694	HIT protein, member of the HIT-family [Methanococcus jannaschii]	61	41	447
171	9	9675	7948	gi1467416	similar to SpoVB [Bacillus subtilis]	61	38	1728
174	3	1042	2140	gi1216374	glutaryl 7-ACA acylase precursor [Bacillus laterosporus]	61	49	1299
190	4	5034	4111	gi1409286	barU [Bacillus subtilis]	61	37	924
216	1	2	190	gi1415861	eukaryotic initiation factor 2 beta (eIF-2 beta) [Cryptosporidium parvum]	61	29	189
227	7	4161	5048	gi1216341	ORF for methionine amino peptidase [Bacillus subtilis]	61	41	888
238	4	1959	3047	gi1409543	[CbrC protein [Erwinia chrysanthemi]	61	38	1089
247	1	2	694	gi1537231	ORF_579 [Escherichia coli]	61	38	693
247	2	678	1034	gi1142226	chvD protein [Agrobacterium tumefaciens]	61	40	357
257	2	3523	2627	gi1699379	glv-1 protein [Mycobacterium leprae]	61	40	897
268	2	3419	3051	gi140364	ORF1 [Clostridium acetobutylicum]	61	41	369
275	4	4821	4827	gi1204848	hypothetical protein (CP:H87049_57) [Haemophilus influenzae]	61	36	207
277	1	1	1845	gi1748897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] p1r[A56390][A56390] mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - Streptococcus pneumoniae	61	45	1845
278	9	8003	7032	gi1467462	cysteine synthetase A [Bacillus subtilis]	61	43	972
278	10	9878	8535	gi11205919	Na ⁺ and Cl ⁻ dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	61	38	1344
283	1	1	366	gi1755607	polyA polymerase [Bacillus subtilis]	61	36	366
288	2	1918	1496	gi1388108	cell wall enzyme [Enterococcus faecalis]	61	43	423
291	1	86	334	gi1454265	FBP3 [Petunia hybrida]	61	38	249
318	1	1104	694	gi1290531	similar to beta-glucoside transport protein [Escherichia coli] sp131451 PT18_ECOLI PTS SYSTEM, ARBUTIN-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT [EC 2.7.1.69]	61	47	411
330	2	1912	1190	gi11001805	hypothetical protein [Synecocystis sp.]	61	41	723

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
385	2	1513	1025	gi 533098	DnaD protein [Bacillus subtilis]	61	42	489
426	1	794	399	gi 1103853	YggF [Bacillus subtilis]	61	44	396
438	3	810	1421	gi 1293660	AbrA2 [Streptomyces coelicolor]	61	36	612
454	1	1580	792	gi 733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellium discoideum]	61	30	789
464	2	784	560	gi 1123120	CSJ87.5 gene product [Caenorhabditis elegans]	61	38	225
470	8	6077	7357	gi 623073	ORF360; putative [Bacteriophage LL-H]	61	47	1281
509	1	554	279	gi 467484	unknown [Bacillus subtilis]	61	45	276
555	3	1916	1296	gi 141800	anthranilate synthase glutamine amidotransferase [Acinetobacter alcoaceticus]	61	42	621
569	1	1711	857	gi 467090	B233_C2_195 [Mycobacterium leprae]	61	47	855
585	2	961	803	sp P36886 SURF_	SURVIVAL PROTEIN SURE HOMOLOG (FRAGMENT)	61	33	159
592	3	1694	1422	gi 1221602	immunity repressor protein [Haemophilus influenzae]	61	32	273
603	1	43	357	gi 507738	Ilmp [Vibrio parahaemolyticus]	61	33	315
669	1	2467	1235	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	61	37	1233
675	3	805	1101	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis]	61	36	297
703	1	1656	829	gi 537181	ORF_470 [Escherichia coli]	61	32	828
728	1	1628	816	gi 806281	DNA polymerase I [Bacillus stearothermophilus]	61	39	813
821	1	61	318	gi 709992	hypothetical protein [Bacillus subtilis]	61	38	258
856	2	2313	1567	gi 609310	portal protein gp3 [Bacteriophage HK97]	61	40	747
923	1	1081	542	gi 143213	putative [Bacillus subtilis]	61	38	540
1124	1	59	370	gi 1107541	C3D9.8 [Caenorhabditis elegans]	61	26	312
1492	1	348	276	gi 406397	unknown [Mycoplasma genitalium]	61	32	273
1602	1	46	318	gi 733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostellium discoideum]	61	34	273
2500	1	577	290	gi 1045964	hypothetical protein [CB:U14001_297] [Mycoplasma genitalium]	61	31	288
2968	1	2	808	gi 397526	clumping factor [Staphylococcus aureus]	61	55	807
3076	1	3	248	gi 149373	ORF 1 [Lactococcus lactis]	61	41	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3609	2	207	401	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearotherophilus</i> [Bacillus subtilis]	61	39	195
3662	1	1477	740	gi 1303813	VggW [Bacillus subtilis]	61	42	738
3672	1	2	442	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	441
3724	1	2	220	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	61	41	219
3728	1	3	398	gi 677943	AppD [Bacillus subtilis]	61	46	396
3884	1	3	401	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	399
3971	1	3	383	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	45	381
4038	1	661	359	gi 1339970	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	61	24	303
4041	1	546	274	gi 413953	lpa-29d gene product [Bacillus subtilis]	61	48	273
4047	1	1	402	gi 528991	unknown [Bacillus subtilis]	61	42	402
4102	1	1	345	gi 976025	HraA [Escherichia coli]	61	46	345
4155	1	1	336	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	336
4268	1	463	233	gi 450688	hdsM gene of Ecoprr1 gene product [Escherichia coli] pir[S38437]S38437 hdsM protein - Escherichia coli pir[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	61	38	231
4374	1	542	273	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	270
4389	1	2	172	gi 147516	ribokinase [Escherichia coli]	61	35	171
4621	1	2	268	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir[A56390]A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	267
4663	1	27	227	gi 976025	HraA [Escherichia coli]	61	50	201
4	6	6663	5536	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearotherophilus</i> [Bacillus subtilis]	60	43	1128

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
11	6	3426	3725	gi 410748	ring-infected erythrocyte surface antigen (Plasmodium falciparum) pfrA2526 A2526 ring-infected erythrocyte surface antigen precursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp P13830 RESA_PLAFF RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN RE	60	24	300
11	14	11035	10313	gi 1217651	carboxyl reductase (NADPH) (Rattus norvegicus)	60	28	723
16	12	11917	12930	gi 1001453	hypothetical protein (Synchocystis sp.)	60	37	1014
33	1	26	469	gi 388109	regulatory protein (Enterococcus faecalis)	60	41	444
37	13	10914	9834	gi 1336656	Orf1 (Bacillus subtilis)	60	40	981
39	4	4364	4522	gi 14872	ORF 4 (Saccharomyces kluyveri)	60	47	159
41	1	2047	1025	gi 1142822	D-alanine racemase cds (Bacillus subtilis)	60	39	1023
43	4	2474	3607	gi 468046	para-nitrobenzyl esterase (Bacillus subtilis)	60	40	1134
44	10	6756	7769	gi 414234	thiF (Escherichia coli)	60	52	1014
45	10	8874	9074	gi 343949	var1(40.0) (Saccharomyces cerevisiae)	60	44	201
56	18	27842	26430	gi 468764	mecR gene product (Rhizobium meliloti)	60	35	1413
60	2	173	388	gi 1303864	YggQ (Bacillus subtilis)	60	33	216
63	2	357	1619	gi 467124	ureD; B229_C3_234 (Mycobacterium leprae)	60	43	1263
69	1	787	395	gi 1518853	OafA (Salmonella typhimurium)	60	36	393
88	1	1	1188	gi 1480429	putative transcriptional regulator (Bacillus stearothermophilus)	60	30	1188
92	6	4735	3881	gi 349227	transmembrane protein (Escherichia coli)	60	37	855
92	7	5996	4923	gi 466613	nlkB (Escherichia coli)	60	38	1074
93	1	949	476	gi 1510925	coenzyme F420-reducing hydrogenase, beta subunit (Methanococcus jannaschii)	60	27	474
96	6	7166	7478	gi 472715	accessory protein (Mycobacterium phaeocola)	60	30	213
98	6	3212	4069	gi 467425	unknown (Bacillus subtilis)	60	42	858
102	10	7158	7430	gi 143092	acetolactate synthase small subunit (Bacillus subtilis) sp P37252 ILVH_BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC 1.3.18) (AIAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS)	60	37	273
109	11	9127	10515	gi 1255259	o-succinylbenzoic acid (OSB) CoA ligase (Staphylococcus aureus)	60	28	1389
109	12	10499	11656	gi 141954	beta-ketothiolase (Alcaligenes eutrophus)	60	41	1158
119	2	4630	3134	gi 1524280	unknown (Mycobacterium tuberculosis)	60	45	1497

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
121	9	6957	7646	gi 1107529	ceuC gene product [Campylobacter coli]	60	35	690
140	7	7704	6013	gi 1146547	kdpA [Escherichia coli]	60	45	1692
145	1	2	703	gi 11460077	unknown [Mycobacterium tuberculosis]	60	23	702
150	3	2809	2216	gi 1146230	putative [Bacillus subtilis]	60	40	594
157	2	1389	961	gi 11303975	YqjX [Bacillus subtilis]	60	30	429
158	5	5125	4769	gi 11449288	unknown [Mycobacterium tuberculosis]	60	36	357
159	1	511	257	gi 1580922	murD gene product [Bacillus subtilis]	60	43	255
160	1	159	1187	gi 1204532	hypothetical protein (GB:L19201.29) [Haemophilus influenzae]	60	34	1029
161	14	8249	7866	gi 11496003	ORF3; PepY; putative oligoendopeptidase based on homology with Lactococcus lactis PepF (GenBank Accession Number Z32522) [Caldicellulosiruptor saccharolyticus]	60	34	384
172	3	1331	2110	gi 1485280	28.2 kDa protein [Streptococcus pneumoniae]	60	33	780
173	2	4082	2460	gi 1524397	glycine betaine transporter OpuD [Bacillus subtilis]	60	41	1623
173	4	5163	4953	gi 1100737	NADP dependent leukotoxin b4 12-hydroxydehydrogenase [Sus scrofa]	60	44	1011
198	1	3	995	gi 413943	ipa-19d gene product [Bacillus subtilis]	60	42	993
201	4	3641	4573	sp P37028 YADT_	HYPOTHETICAL 29.4 KD PROTEIN IN HENL-PPS INTERGENIC REGION PRECURSOR.	60	37	933
203	1	3269	2415	gi 1927798	D0719.34n; CA1; 0.14 [Saccharomyces cerevisiae]	60	43	855
206	9	12234	12515	sp P37347 YECD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASP5 5'-REGION.	60	47	282
212	4	1213	1410	gi 332711	hemagglutinin-neuraminidase fusion protein [Human parainfluenza virus 3]	60	34	198
214	1	65	1153	gi 1204366	hypothetical protein (GB:U14001.130) [Haemophilus influenzae]	60	36	1089
237	1	2	937	gi 149377	Hsd [Lactococcus lactis]	60	40	936
241	6	5696	4998	gi 1046160	hypothetical protein (GB:U00021.5) [Mycoplasma genitalium]	60	37	699
260	6	5919	6485	gi 1431950	similar to a B.subtilis gene (GB: BACHEMY_5) [Clostridium asteurianum]	60	35	567
264	1	2432	1218	gi 197526	clumping factor [Staphylococcus aureus]	60	53	1215
267	1	3	1409	gi 148316	NAH-antiporter protein [Enterococcus hirae]	60	27	1407
275	3	3804	4595	pir P36889 P368	leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis [strain IL1403]	60	35	792
291	3	860	1198	gi 1208889	coded for by C. elegans cDNA yk130a12.5; contains C2H2-type zinc fingers [Caenorhabditis elegans]	60	33	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
307	6	3421	3176	gi11070014	protein-dependent [Bacillus subtilis]	60	36	246
316	8	4957	5823	gi1413952	ppe-26d gene product [Bacillus subtilis]	60	41	867
328	4	2996	3484	gi11204484	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	gi11205449	colicin V production protein (pur regulon) [Haemophilus influenzae]	60	37	525
357	1	1062	532	gi1487842	single-stranded DNA-specific exonuclease [Escherichia coli]	60	41	531
375	2	96	362	gi11057	adenylyl cyclase gene product [Saccharomyces kluyveri] t1201145[OYBYK]	60	47	267
					adenylyl cyclase [EC 4.6.1.1] - yeast ccharomyces kluyveri			
397	1	66	416	gi1709999	glucarate dehydratase [Bacillus subtilis]	60	37	351
409	1	2	163	gi1499700	glycogen phosphorylase [Saccharomyces cerevisiae]	60	35	162
453	4	916	1237	gi11196899	unknown protein [Staphylococcus aureus]	60	36	324
453	7	3838	3620	sp12222[VCFL_	HYPOTHEICAL 226 KD PROTEIN (ORF 1901).	60	31	219
470	2	622	945	pir1530782[S307	integrin homolog - yeast [Saccharomyces cerevisiae]	60	31	324
500	1	118	606	gi1467407	unknown [Bacillus subtilis]	60	36	489
503	3	752	982	gi1167835	lysozin heavy chain [Dictyostelium discoideum]	60	34	231
505	4	2238	3563	gi11510732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	gi1143331	alkaline phosphatase regulatory protein [Bacillus subtilis]	60	41	1041
					pir1427650[regulatory protein phor - Bacillus subtilis]			
					sp123545[PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN H0R (EC 2.7.3.-)]			
543	1	1	465	gi1311103	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	60	40	465
545	1	1	726	gi11498192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	gi11477402	tex gene product [Bordetella pertussis]	60	42	1053
578	1	974	489	gi11205129	H. influenzae predicted coding region H10882 [Haemophilus influenzae]	60	42	486
594	1	1	624	gi11212755	adenylyl cyclase [Aeromonas hydrophila]	60	45	624
604	1	3	530	gi1145925	[fecB [Escherichia coli]	60	42	528
620	1	926	465	gi11205483	bicyclic resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	gi11486242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	gi11205136	serine hydroxymethyltransferase (serine methylase) [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
684	1	1082	843	gi 1205538	hypothetical protein [G8:U14003.302] [Haemophilus influenzae]	60	39	240
786	1	967	485	gi 1402944	orfH1 gene product [Bacillus subtilis]	60	46	483
844	1	588	346	gi 790943	urea amidolyase [Bacillus subtilis]	60	40	243
851	1	1	726	gi 159661	GMP reductase [Ascaris lumbricoides]	60	41	726
871	1	1746	874	gi 1001493	hypothetical protein [Synecococcus sp.]	60	39	873
896	1	1558	839	gi 604926	NADH dehydrogenase, subunit 5 [Schizosaccharomyces pombe]	60	39	720
908	2	448	753	gi 662880	novel hemolytic factor [Bacillus cereus]	60	31	306
979	1	2	595	gi 1429255	putative; orf1 [Bacillus subtilis]	60	30	594
1078	1	669	502	gi 581055	inner membrane copper tolerance protein [Escherichia coli] gi 871029	60	40	168
					disulphide isomerase like protein [Escherichia coli] pir S47295 S47295			
					inner membrane copper tolerance protein - escherichia coli			
1112	1	1150	620	gi 407885	ORF3 [Streptomyces griseus]	60	34	531
1135	1	484	275	gi 1171407	Vps8p [Saccharomyces cerevisiae]	60	36	210
1146	1	17	562	gi 1239981	hypothetical protein [Bacillus subtilis]	60	36	546
1291	1	716	360	pir S57510 S575	carboxyl esterase - Acinetobacter calcoaceticus	60	30	357
1332	1	336	169	gi 1222056	aminotransferase [Haemophilus influenzae]	60	44	168
1429	1	3	146	gi 1205619	ferritin like protein [Haemophilus influenzae]	60	39	144
1722	1	570	286	gi 240052	dlhydroflavonol-4-reductase, DFR [Hordeum vulgare-barley, cv. Gula, optide, 354 aa]	60	36	285
2350	1	385	200	gi 497626	ORF 1 [Plasmid pAQ1]	60	20	186
2936	1	519	310	gi 508981	prephenate dehydratase [Bacillus subtilis]	60	48	210
3027	1	568	302	gi 1146199	putative [Bacillus subtilis]	60	37	267
3084	1	20	208	gi 140784	orf-1; novel antigen [Staphylococcus aureus]	60	51	189
3155	1	2	226	gi 1046097	cytadherence-accessory protein [Mycoplasma genitalium]	60	34	225
3603	1	368	186	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA ahydrogenase alpha-subunit [Rattus norvegicus]	60	42	183
3665	1	486	244	gi 151259	HMG-CoA reductase [EC 1.1.1.88] [Pseudomonas mevalonii] pir A44756 A44756	60	42	243
					hydroxymethylglutaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.			
3747	1	3	146	gi 474192	lucC gene product [Escherichia coli]	60	36	144

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match junction	match gene name	% sim	% ident	length (nt)
3912	1	3	335	gi 488695	novel antigen: orf-2 [Staphylococcus aureus]	60	44	333
4072	1	3	272	gi 405879	yleH [Escherichia coli]	60	33	270
4134	1	510	352	gi 780656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae] gi 780656 chemoreceptor protein [Rhizobium leguminosarum bv. lotiae]	60	28	159
4207	2	677	402	gi 602031	similar to trimethylamine DH [Mycoplasma capricolum] pir S49950 S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SC3) (fragment)	60	41	276
4243	1	127	324	gi 899317	peptide synthetase module [Microcystis aeruginosa] pir S49111 S49111 probable amino acid activating domain - Microcystis aeruginosa (fragment) (SUB 144-528)	60	42	198
4310	1	624	313	gi 508980	pheB [Bacillus subtilis]	60	28	312
4345	1	343	173	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA elhydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
4382	1	498	280	gi 47382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	60	48	219
4474	1	53	223	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA elhydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	3523	gi 426446	VibB protein [Salmonella typhi]	59	39	996
33	2	707	1483	gi 548604 S486	hypothetical protein - Mycoplasma capricolum (SC3) (fragment)	59	33	777
33	5	4651	5853	gi 6721	P5982.3 [Caenorhabditis elegans]	59	33	1203
17	2	3228	2299	gi 142833	ORF2 [Bacillus subtilis]	59	37	910
38	21	16784	16593	gi 912576	BIP [Pseudocystium tricornutum]	59	40	192
52	3	2648	2349	gi 536972	ORF_030a [Escherichia coli]	59	44	300
54	12	14181	13402	gi 483940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi 508176	Gat-1-P-DH, NAD dependent [Escherichia coli]	59	40	1059
66	1	986	495	gi 1303901	Yqhr [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi 912461	nlkC [Escherichia coli]	59	37	909
70	7	5383	6366	gi 1399822	PhoD precursor [Rhizobium meliloti]	59	46	984
78	1	1	1449	gi 971345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 VME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION.	59	39	1449
82	10	14329	15534	gi 490328	LORF P (unidentified)	59	44	1206

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
89	2	1602	958	gi1642801	unknown [Saccharomyces cerevisiae]	59	32	645
96	4	4940	5473	gi1333802	protein of unknown function [Rhodobacter capsulatus]	59	33	534
98	1	2	820	gi1467421	similar to B. subtilis DnaH [Bacillus subtilis]	59	34	819
119	1	166	1557	gi1143122	ORF B; putative [Bacillus firmus]	59	36	1392
120	10	6214	6756	gi115354	ORF 55.9 [Bacteriophage T4]	59	39	543
120	16	112476	13510	gi11086575	BatA [Rhizobium meliloti]	59	44	1035
133	1	386	195	gi1984737	catalase [Campylobacter jejuni]	59	38	192
130	1	370	645	gi1256634	25-8% identity over 120 aa with the Synenococcus sp. MpeV protein; putative [Bacillus subtilis]	59	31	276
131	4	5278	5712	gi11510655	hypothetical protein (SP:P42297) [Methanococcus jannaschii]	59	39	435
164	1	3	509	gi11001342	hypothetical protein [Synecocystis sp.]	59	41	507
164	4	1529	2821	gi11205165	hypothetical protein (SP:P37764) [Haemophilus influenzae]	59	35	1293
164	19	19643	21376	gi11001381	hypothetical protein [Synecocystis sp.]	59	34	1734
173	3	4727	3717	gi11184121	leuxin-induced protein [Vigna radiata]	59	50	1011
179	2	2218	1688	gi1143036	unidentified gene product [Bacillus subtilis]	59	33	531
195	12	12669	11503	gi1762778	hfs gene product [Anabaena azollae]	59	41	1167
201	5	4702	5670	gi11510240	hemin permease [Methanococcus jannaschii]	59	32	969
201	7	5719	6315	gi11511456	M. jannaschii predicted coding region M1437 [Methanococcus jannaschii]	59	34	597
209	1	102	461	gi1204566	hypothetical protein (GB:X73124.53) [Haemophilus influenzae]	59	42	360
214	3	1050	2234	gi1551531	2-nitropropane dioxygenase [Miliopsis saturnus]	59	36	1185
214	5	3291	4135	gi11303709	YrkJ [Bacillus subtilis]	59	32	843
217	2	3381	2167	gi1290489	dcp (CG Site No. 18430) [Escherichia coli]	59	44	1215
237	5	3078	3785	gi1149382	hflBA [Lactococcus lactis]	59	38	708
251	2	376	960	gi11303791	YqjJ [Bacillus subtilis]	59	34	585
286	1	1621	812	gi1146551	transmembrane protein (kdp) [Escherichia coli]	59	31	810
316	5	4978	3860	gi1405879	yeiH [Escherichia coli]	59	32	1119
370	3	600	761	gi11303794	YqeM [Bacillus subtilis]	59	35	162

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% idunt	length (nt)
382	1	1009	506	gi 347513	orf3 [Haemophilus influenzae]	59	34	504
391	3	1620	1273	gi 152901	ORF 3 [Spirochaeta aurantiola]	59	37	348
406	3	2805	1705	gi 709992	hypothetical protein [Bacillus subtilis]	59	34	1101
426	5	3802	3245	gi 1204610	ironIII diolate transport ATP-binding protein PECE [Haemophilus influenzae]	59	36	558
429	2	1513	1148	gi 1064809	homologous to sp.NTRA_ECOLI [Bacillus subtilis]	59	42	366
460	2	708	1301	gi 466882	ppa1: BA496_C2_189 [Mycobacterium leprae]	59	37	594
461	4	2212	3135	gi 1498295	homoserine kinase homolog [Streptococcus pneumoniae]	59	37	924
473	1	2929	1607	gi 147989	trigger factor [Escherichia coli]	59	40	1323
480	8	5862	6110	gi 1205311	(3R)-hydroxymyristol acyl carrier protein dehydrase [Haemophilus influenzae]	59	40	249
521	1	14	1354	pir A25620 A256	staphylocoagulase - Staphylococcus aureus (fragment)	59	32	1341
534	4	2994	4073	gi 153746	mannitol-phosphate dehydrogenase [Streptococcus mutans] pir C44798 C44798	59	36	1080
535	1	1	954	gi 1469939	group B oligopeptidase PepB [Streptococcus agalactiae]	59	33	954
551	3	2836	3186	gi 1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	59	45	351
573	2	449	940	gi 386581	ORF VAL022 [Saccharomyces cerevisiae]	59	36	492
650	1	5	748	gi 396400	similar to eukaryotic Na+/H+ exchanger [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOAR-ACS HYPERGENIC REGION (0549)	59	30	744
664	1	566	285	gi 1262748	LukP-PV like component [Staphylococcus aureus]	59	33	282
670	1	3	455	gi 1122758	unknown [Bacillus subtilis]	59	42	453
674	3	543	929	gi 293033	Integrase [Bacteriophage phi-LC3]	59	46	387
758	1	349	176	gi 1500472	M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii]	59	37	174
771	2	2270	1461	gi 522150	bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp P3912 BPAL_STRAU NON-ILAEH BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE (BPO1), (SUB 2-275)	59	44	810
825	1	2191	1097	gi 397526	clumping factor [Staphylococcus aureus]	59	47	1095
1052	2	1094	723	gi 289262	comE ORF3 [Bacillus subtilis]	59	36	372
1152	1	373	188	gi 1127668	ORF238 gene product [Porphyra purpurea]	59	37	186

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1198	1	492	247	gi1142439	ATP-dependent nuclease [Bacillus subtilis]	59	26	246
1441	1	468	235	gi11045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	59	37	234
2103	1	1	186	gi1459250	triacylglycerol lipase [Galactomyces geotrichum]	59	33	186
2205	1	793	398	gi11303794	Yqem [Bacillus subtilis]	59	38	396
2578	1	484	284	gi1258003	insulin-like growth factor binding protein complex acid-labile subunit [rats, liver, Peptide, 603 aa]	59	48	201
2967	2	145	348	gi11212730	YqhK [Bacillus subtilis]	59	44	204
3012	1	3	248	gi1773571	neurofilament protein NF70 [Helix aspersa]	59	31	246
3544	1	3	401	gi11055218	crotonase [Clostridium acetobutylicum]	59	42	399
3548	1	3	401	gi11055218	crotonase [Clostridium acetobutylicum]	59	42	399
3580	1	698	351	gi11055218	crotonase [Clostridium acetobutylicum]	59	42	348
3720	1	722	363	gi11408494	homologous to penicillin acylase [Bacillus subtilis]	59	36	360
4171	1	3	296	gi11055218	crotonase [Clostridium acetobutylicum]	59	42	294
4305	1	618	310	gi11524193	unknown [Mycobacterium tuberculosis]	59	39	309
18	1	1242	622	gi1146913	N-acetylglucosamine transport protein [Escherichia coli] p1c[B29895]WQEC2W phosphotransferase system enzyme II (EC 7.1.69), N-acetylglucosamine-specific - Escherichia coli sp P09323 PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IABC COMPONENT (EITA	58	43	621
20	7	7020	5845	gi150502	collagen alpha chain precursor (AA -27 to 1127) [Mus musculus]	58	50	1176
21	5	3234	3626	gi11054860	phosphoribosyl anthranilate isomerase [Thermotoga maritima]	58	32	393
23	2	2841	1669	gi11276880	Epac [Streptococcus thermophilus]	58	29	1173
23	10	9101	8090	gi1131133 A311	diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa	58	37	1212
38	29	22555	22884	gi1973249	vestitone reductase [Medicago sativa]	58	37	330
44	1	2	406	gi1289272	ferriochrome-binding protein [Bacillus subtilis]	58	33	405
45	1	1	552	gi129464	embryonic myosin heavy chain (1085 AA) [Homo sapiens] tr S12460 S12460 myosin beta heavy chain - human	58	33	552
55	2	759	538	gi1158852	glucose regulated protein [Echinococcus multilocularis]	58	32	222
62	13	8493	8068	gi1975353	kinase-associated protein B [Bacillus subtilis]	58	35	426
63	3	1553	1717	gi1166926	[Arabidopsis thaliana unidentified mRNA sequence, complete cds.], one product [Arabidopsis thaliana]	58	35	165

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
67	13	11229	11229	gi 1228083	NADH dehydrogenase subunit 2 (Chorthippus parallelus)	58	41	789
96	8	8208	9167	gi 709992	hypothetical protein [Bacillus subtilis]	58	42	960
107	2	2065	1364	gi 806327	Escherichia coli hrpA gene for A protein similar to yeast PRP16 and PRP22 [Escherichia coli]	58	37	702
112	7	4519	5613	gi 155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir[A42289]	58	38	1095
114	6	7318	6503	gi 1377843	glucose-fructose oxidoreductase (EC 1.1.1.-) precursor - Zymomonas mobilis [unknown [Bacillus subtilis]]	58	38	816
143	2	2261	1395	pir[A45605]A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	58	31	867
151	2	717	950	gi 3370261	[unknown [Mycobacterium tuberculosis]]	58	31	234
156	6	6015	4627	gi 1209277	pCTH01 gene product [Chlamydia trachomatis]	58	41	1389
154	16	14281	13541	gi 146613	DNA ligase (EC 6.5.1.2) [Escherichia coli]	58	39	741
155	3	2269	1892	gi 1303917	YqjB [Bacillus subtilis]	58	34	378
174	1	1056	539	gi 304198	hypothetical protein [Bacillus subtilis]	58	26	528
189	4	1533	1769	gi 467383	DNA binding protein (probable) [Bacillus subtilis]	58	25	237
201	3	2669	3307	gi 1511453	endonuclease III [Methanococcus jannaschii]	58	34	639
208	1	2	238	gi 1276729	phycobilisome linker polypeptide [Porphyra purpurea]	58	29	237
220	11	14575	13058	gi 397526	clumping factor [Staphylococcus aureus]	58	51	1518
231	3	1629	1474	gi 1002520	MutS [Bacillus subtilis]	58	45	156
233	6	4201	3497	gi 1463023	No definition line found [Caenorhabditis elegans]	58	39	705
243	10	9303	10082	gi 537207	ORF_277 [Escherichia coli]	58	32	780
257	1	331	1143	gi 1340128	ORF1 [Staphylococcus aureus]	58	44	813
302	2	460	801	gi 40174	ORF X [Bacillus subtilis]	58	34	342
307	31	6984	6127	gi 1303842	YqjU [Bacillus subtilis]	58	30	858
321	3	1914	2747	gi 1239996	hypothetical protein [Bacillus subtilis]	58	41	834
342	4	2724	3497	gi 1454838	ORF 6: putative [Pseudomonas aeruginosa]	58	41	774
348	1	1	663	gi 467478	[unknown [Bacillus subtilis]]	58	36	663
401	2	384	605	gi 143407	[para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]]	58	53	222

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	naçp gene name	% sim	% ident	length (nt)
437	1	325	1554	gi1301866	YngS (Bacillus subtilis)	58	35	1230
445	1	105	1442	gi1581383	protein A (Staphylococcus aureus)	58	32	1318
453	3	789	965	gi1009455	unknown (Schizosaccharomyces pombe)	58	34	177
453	5	2748	2047	gi1537214	yjgC gene product (Escherichia coli)	58	40	702
479	2	731	1444	gi1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative (Bacillus subtilis)	58	36	714
490	1	909	547	gi1580920	codB (gtaA) polypeptide (AA 1-673) (Bacillus subtilis) pir[S06048/S06048 probable rodd protein - Bacillus subtilis sp]P13484/TAGE BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHN-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOLIC ACID BIOSYNTHESIS ROBIN E)	58	36	363
517	1	1	1164	sp147264/Y018_	HYPOTHETICAL HELICASE M0018	58	30	1164
517	6	4182	4544	gi1453422	orf268 gene product (Hycoplasma hominis)	58	29	363
546	3	2802	4019	gi1886052	restriction modification system S subunit (Spiroplasma citri) gi1886052 restriction modification system S subunit (Spiroplasma citri)	58	37	1218
562	1	3	179	gi143831	nifS protein (AA 1-400) (Klebsiella pneumoniae)	58	34	177
600	2	1347	1156	gi1181819	unknown (Pseudomonas aeruginosa)	58	48	192
604	2	1231	1001	gi1001353	hypothetical protein (Synecocystis sp.)	58	41	231
619	1	1	504	gi1903748	Integral membrane protein (Homo sapiens)	58	43	504
625	1	2	364	gi1208474	hypothetical protein (Synecocystis sp.)	58	43	363
635	1	1492	755	gi1151095	transaldolase (Methanococcus jannaschii)	58	41	738
645	1	1	846	gi1677882	ileal sodium-dependent bile acid transporter (Rattus norvegicus) gi1677882 ileal sodium-dependent bile acid transporter (Rattus norvegicus)	58	33	846
645	3	906	1556	gi1239999	hypothetical protein (Bacillus subtilis)	58	41	651
665	1	771	532	gi1204262	hypothetical protein (GB:110128_61) (Haemophilus influenzae)	58	39	240
674	1	635	327	gi1498817	ORF8; homologous to small subunit of phage terminases (Bacillus subtilis)	58	39	309
675	2	1312	806	gi142181	osmC gene product (Escherichia coli)	58	28	507
745	1	618	310	gi1205432	coenzyme PQQ synthetase protein III (pqgIII) (Haemophilus influenzae)	58	32	309
799	2	242	1174	gi1204669	collagenase (Haemophilus influenzae)	58	36	933
800	2	1096	614	gi1171963	trna isopentenyl transferase (Saccharomyces cerevisiae) sp167884/M05_YEAST TRNA ISOPENTENYLTRANSFERASE (EC 2.3.1.8) ISOPENTENYL-DIPHOSPHATE, TRNA ISOPENTENYLTRANSFERASE (IPP TRANSFERASE) (IPPT)	58	37	483

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
854	1	1108	605	gi1466778	lysine specific peptidase [Escherichia coli]	58	44	504
865	1	481	242	gi1861199	protoporphyrin IX Mg-chelatase subunit precursor [Hordeum vulgare]	58	33	240
891	1	3	527	gi1293660	AbsA2 [Streptomyces coelicolor]	58	31	525
942	1	931	467	gi1405567	traw [plasmid pSK41]	58	30	465
1002	1	952	521	gi1577649	preLUXK [Staphylococcus aureus]	58	34	432
1438	1	1	261	gi1581558	isoleucyl tRNA synthetase [Staphylococcus aureus] sp141368[SYIP_STAAU ISOLEUCYL-TRNA SYNTHETASE, MUPIROCIN RESISTANT EC 6.1.1.5] (ISOLEUCINE--TRNA LIGASE) (ILERS) (MUPIROCIN RESISTANCE PROTEIN)	58	30	261
1442	1	2	463	gi1971394	similar to Acc-No. D36185 [Escherichia coli]	58	34	462
1873	1	480	241	gi1133951	small subunit of NADH-dependent glutamate synthase [Plectonoma boryanum]	58	38	240
1876	1	3	158	gi1529216	No definition line found [Caenorhabditis elegans] sp146650[YLA7_CAEEL HYPOTHETICAL 7.3 KD PROTEIN F23F12.7 IN HR05086 III]	58	33	156
1989	1	108	401	gi1405458	YneR [Bacillus subtilis]	58	29	294
2109	1	3	401	gi11001801	hypothetical protein [Synecocystis sp.]	58	31	399
2473	1	288	145	gi1510140	ligandopeptidase F [Lactococcus lactis]	58	38	144
2523	1	452	228	gi1644873	catabolic dehydroquinase dehydratase [Acinetobacter calcoaceticus]	58	37	225
3041	1	2	211	gi1205367	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	58	39	210
3094	1	3	263	gi1185288	isochorismate synthase [Bacillus subtilis]	58	38	261
3706	1	3	383	gi1456614	mevalonate kinase [Arabidopsis thaliana]	58	48	381
3854	1	1	402	gi1808869	human gcp372 [Homo sapiens]	58	32	402
4082	1	51	224	gi1508551	ribulose-1,5 bisphosphate carboxylase large subunit -methyltransferase [Pisum sativum]	58	37	174
4278	1	3	206	gi1180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi1182737 cerebellar degeneration-associated protein [Homo sapiens] p1c[A29770[A29770 cerebellar degeneration-related protein - human	58	37	204
19	7	7818	7363	gi11001516	hypothetical protein [Synecocystis sp.]	57	31	456
23	11	9663	8872	gi1606066	ORF_f256 [Escherichia coli]	57	29	792
31	1	4801	2402	gi1531146	ORF3 [Streptomyces coelicolor]	57	32	2400
38	14	11611	10796	gi1144859	ORF B [Clostridium perfringens]	57	31	816
46	14	11263	13046	gi11001319	hypothetical protein [Synecocystis sp.]	57	25	984

TABLE 2

5 aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
51	3	1411	1187	pir B33856 B338	hypothetical 60K protein - <i>Bacillus ephaericus</i>	57	38	225
54	1	1	453	gi 684950	staphylococcal accessory regulator A (<i>Staphylococcus aureus</i>)	57	31	453
75	1	3	239	gi 1000470	C2787.7 (<i>Caenorhabditis elegans</i>)	57	42	237
92	5	3855	3061	gi 143607	sporulation protein (<i>Bacillus subtilis</i>)	57	35	795
96	3	4006	4773	gi 144297	acetyl esterase (XycC) (<i>Caldocellum saccharolyticum</i>) pir B37202 B37202 acetyl esterase (EC 3.1.1.6) (XycC) - <i>Caldocellum saccharolyticum</i>	57	34	768
107	3	1480	2076	gi 460955	TagE (<i>Vibrio cholerae</i>)	57	42	597
109	8	5340	5933	gi 1438846	Unknown (<i>Bacillus subtilis</i>)	57	41	594
112	9	6679	7701	gi 1486250	Unknown (<i>Bacillus subtilis</i>)	57	33	1023
114	4	6384	4108	gi 871456	putative alpha subunit of formate dehydrogenase (<i>Methanobacterium thermoautotrophicum</i>)	57	37	2277
126	2	430	1053	gi 288301	ORF2 gene product (<i>Bacillus megaterium</i>)	57	37	624
131	5	6537	6277	gi 1511160	M. jannaschii predicted coding region M31163 (<i>Methanococcus jannaschii</i>)	57	38	261
133	3	2668	2201	gi 1303912	YqhW (<i>Bacillus subtilis</i>)	57	40	468
133	4	3383	2784	gi 1221884	(furea?) amidolyase (<i>Haemophilus influenzae</i>)	57	37	600
147	4	2164	1694	gi 467469	Unknown (<i>Bacillus subtilis</i>)	57	33	471
160	2	1293	1060	gi 558604	chitin synthase 2 (<i>Neurospora crassa</i>)	57	28	234
163	8	5687	4764	gi 145580	farD gene product (<i>Escherichia coli</i>)	57	38	924
168	6	4336	5325	gi 39782	33kDa lipoprotein (<i>Bacillus subtilis</i>)	57	32	990
170	5	3297	3455	gi 603404	Yer164p (<i>Saccharomyces cerevisiae</i>)	57	37	159
221	6	8026	6809	gi 1136221	carboxypeptidase (<i>Sulfolobus solfataricus</i>)	57	32	1218
228	3	1348	1791	gi 288969	fibronectin binding protein (<i>Streptococcus dysgalactiae</i>) pir S33850 S33850 fibronectin-binding protein - <i>Streptococcus dysgalactiae</i>	57	32	444
263	4	4411	3686	gi 1185002	dihydrodipicolinate reductase (<i>Pseudomonas syringae</i> pv. <i>tabaci</i>)	57	42	726
276	1	494	255	gi 396380	No definition line found (<i>Escherichia coli</i>)	57	40	240
283	2	335	1326	gi 773349	SlrA protein (<i>Bacillus subtilis</i>)	57	32	990
297	1	469	236	gi 1334820	reading frame V (<i>Cauliflower mosaic virus</i>)	57	46	234
342	3	1993	2805	gi 1204431	hypothetical protein (GP-P3644) (<i>Haemophilus influenzae</i>)	57	35	813

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
375	6	3340	3741	gi 385177	cell division protein (Bacillus subtilis)	57	26	402
433	6	3286	4011	gi 1524117	alpha-acetolactate decarboxylase (Lactococcus lactis)	57	40	726
470	3	903	1145	gi R04819	protein serine/threonine kinase (Toxoplasma gondii)	57	30	243
487	5	1391	1723	gi 507323	ORF1 (Bacillus steatothermophilus)	57	28	333
498	1	274	852	gi 1334549	NADH-ubiquinone oxidoreductase subunit 4L (Podospora anserina)	57	34	579
503	1	343	173	gi 1502283	organic cation transporter OCT2 (Rattus norvegicus)	57	30	171
505	2	1619	1284	gi 466884	bls96_C2_194 (Mycobacterium leprae)	57	40	336
519	2	1182	2549	gi 1303707	YrkH (Bacillus subtilis)	57	34	1368
522	2	3234	1945	gi 1064809	homologous to spiRPA_ECOLI (Bacillus subtilis)	57	36	1290
538	2	909	1415	gi 153179	phosphorinothricin n-acetyltransferase (Streptomyces coelicolor) pfrJH0246/JH0246 phosphinothricin N-acetyltransferase (EC 2.3.1.-)	57	40	507
547	1	968	486	gi 467340	unknown (Bacillus subtilis)	57	50	483
599	1	1062	532	sp P20692 TYRA_1	PREPIENATE DEHYDROGENASE (EC 1.3.1.121) (PDIH)	57	41	531
620	2	757	572	gi 1107894	unknown (Schizosaccharomyces pombe)	57	38	186
622	2	1600	1130	gi 1173028	thioredoxin II (Saccharomyces cerevisiae)	57	39	471
625	2	362	1114	gi 1262366	hypothetical protein (Mycobacterium leprae)	57	34	753
680	1	1	204	gi 143544	RNA polymerase sigma-30 factor (Bacillus subtilis) pfrJH0246/JH0246 transcription initiation factor sigma H - acillus subtilis	57	30	204
690	1	3	629	gi 466520	pocr (Salmonella typhimurium)	57	29	627
696	1	2	433	gi 413972	ipe-48r gene product (Bacillus subtilis)	57	33	432
704	1	36	638	gi 1499331	M. jannaschii predicted coding region MJ1083 (Methanococcus jannaschii)	57	36	603
732	1	2316	1621	gi 1418999	orf4 (Lactobacillus sake)	57	37	696
746	1	451	227	gi 392973	Rah3 (Aplysia californica)	57	42	225
757	1	20	466	gi 43979	L. curvatus small cryptic plasmid gene for rep protein (Lactobacillus rvaus)	57	45	447
862	1	2	295	gi 1303827	Yqf1 (Bacillus subtilis)	57	21	294
1049	1	907	455	gi 1510108	ORF-1 (Agrobacterium tumefaciens)	57	35	453
1117	1	1387	695	gi 896286	MH2 terminus uncertain (Leishmania tarentolae)	57	28	693

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1136	1	2	322	gi 1303853	Yqf (Bacillus subtilis)	57	38	321
1144	2	1033	611	gi 310083	voltage-activated calcium channel alpha-1 subunit (Rattus orvegicus)	57	46	423
1172	1	1472	738	gi 1511146	M. jannaschii predicted coding region MJ1143 (Methanococcus jannaschii)	57	28	735
1500	2	746	558	gi 142780	putative membrane protein; putative (Bacillus subtilis)	57	35	189
1676	1	659	399	gi 313777	luciferase (Escherichia coli)	57	31	261
2481	1	2	400	gi 1237015	ORF4 (Bacillus subtilis)	57	23	399
3099	1	3	230	gi 1204510	isochlorinate synthase (Haemophilus influenzae)	57	19	228
3122	1	360	181	gi 882472	ORF 0464 (Escherichia coli)	57	40	180
3560	1	2	361	gi 153490	tetracycline C resistance and export protein (Streptomyces laueae)	57	37	360
3850	1	856	434	gi 155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pIR/A42289	57	40	423
3931	1	704	354	gi 413953	lipa-29d gene product (Bacillus subtilis)	57	36	351
3993	1	1	384	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas nevaloni) pIR/A44756	57	39	384
4065	1	793	398	gi 1000365	hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	57	31	396
4100	1	596	300	gi 1086633	nitrate reductase (EC 1.7.99.4) alpha chain - Escherichia coli	57	47	297
4163	1	571	287	gi 21512	T06C10.5 gene product (Caenorhabditis elegans)	57	50	285
4267	2	631	335	gi 1000365	potatin (Solanum tuberosum)	57	38	297
4358	1	3	302	gi 398032	SpolIAG (Bacillus subtilis)	57	32	300
4389	2	108	290	gi 405894	EF (Streptococcus suis)	57	37	183
4399	1	2	232	gi 1483603	1-phosphofructokinase (Escherichia coli)	57	35	231
4481	1	572	288	gi 405879	Pristinamycin I synthase I (Streptomyces pristinaespiralis)	57	44	285
4486	1	512	258	gi 515938	lyeH (Escherichia coli)	57	42	255
4510	1	481	242	gi 1205301	glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	57	38	240
4617	1	468	256	gi 1511222	leukotoxin secretion ATP-binding protein (Haemophilus influenzae)	57	35	213
4	11	12201	11524	gi 149204	restriction modification enzyme, subunit M1 (Methanococcus jannaschii)	56	31	678
					histidine utilization repressor G (Klebsiella aerogenes) pIR/A36730			
					hutG protein - Klebsiella pneumoniae (fragment) sp P19452 HUTG_KLEA			
					FORMINOGLOUTAMATE (EC 3.5.3.8) FORMINOGLOUTAMATE HYDROLASE (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	8	4248	5177	gi11322222	RACH1 [Homo sapiens]	56	33	930
38	28	21179	22264	gi11480705	lipote-protein lipase [Mycoplasma capricolum]	56	34	1086
44	3	1861	2421	gi1490320	Y gene product (unidentified)	56	31	561
44	15	10103	10806	gi1205059	hypothetical protein (GB:U19201.1) [Haemophilus influenzae]	56	39	504
50	6	4820	5161	gi1209931	fiber protein [Human adenovirus type 5]	56	48	342
53	4	2076	2972	gi1623476	transcriptional activator (Providencia stuartii) sp P43463 AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP.	56	30	897
67	6	5656	6594	gi1466613	hns [Escherichia coli]	56	32	939
89	3	2364	1810	gi1482922	protein with homology to pail repressor of B. subtilis [Lactobacillus elbrueckii]	56	39	555
96	1	203	913	gi1145594	cAMP receptor protein (crp) [Escherichia coli]	56	35	711
109	21	118250	117846	gi11204367	hypothetical protein (GB:U14003.278) [Haemophilus influenzae]	56	27	405
112	8	5611	6678	gi1155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir A42289 A42289 glucose-fructose oxidoreductase (EC 1.1.1.-) reductor - Zymomonas mobilis	56	40	1088
131	3	6404	5100	gi1619724	MgtE [Bacillus firmus]	56	30	1305
138	2	65	232	gi1413948	lipa-24d gene product [Bacillus subtilis]	56	31	168
138	4	823	1521	gi1580868	lipa-22r gene product [Bacillus subtilis]	56	31	699
146	2	740	447	gi11046009	M. genitalium predicted coding region MG309 [Mycoplasma genitalium]	56	37	294
149	2	1639	1067	gi11945380	terminase small subunit [Bacteriophage LU-II]	56	35	573
163	1	2	223	gi1143947	glutamine synthetase [Bacteroides fragilis]	56	30	222
166	5	6745	6449	gi1405792	ORF154 [Pseudomonas putida]	56	26	297
187	1	31	393	gi1311237	H(+)-transporting ATP synthase [Zea mays]	56	30	363
190	1	2	373	gi1109686	Prox [Bacillus subtilis]	56	35	372
191	8	11538	9943	gi11581070	acyl coenzyme A synthetase [Escherichia coli]	56	35	1596
195	3	1291	647	gi11510242	collagenase [Methanococcus jannaschii]	56	34	645
230	3	2323	2072	gi1403163	heat shock protein (Clostridium acetobutylicum)	56	39	252
238	5	3383	3775	gi11477533	earA [Staphylococcus aureus]	56	31	393
270	2	813	1712	gi1765073	autolysin [Staphylococcus aureus]	56	41	900

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
290	1	3221	1632	gi1547513	[orf3] [Haemophilus influenzae]	56	34	1590
297	5	1140	1373	gi1511556	[M. jannaschii predicted coding region MJ1561 [Methanococcus jannaschii]	56	40	234
321	2	2947	1799	gi1001801	[hypothetical protein [Synecocystis sp.]	56	31	1149
359	2	1279	641	gi146336	[noli gene product [Rhizobium meliloti]	56	26	639
371	2	360	1823	gi145304	[L-ribulokinase [Escherichia coli]	56	39	1464
391	4	1762	2409	gi11001634	[hypothetical protein [Synecocystis sp.]	56	34	648
402	1	380	192	gi1438904	[5-HT4L receptor [Homo sapiens]	56	48	189
416	4	2480	2109	gi11408486	[HS74A gene product [Bacillus subtilis]	56	31	372
424	3	1756	2334	gi1142471	[acetolactate decarboxylase [Bacillus subtilis]	56	32	579
457	1	1907	1017	gi1205194	[formamidopyrimidine-DNA glycosylase [Haemophilus influenzae]	56	36	891
458	2	2423	1812	gi115466	[terminase [Bacteriophage SP1]	56	37	612
504	2	2152	1283	gi1142681	[lpp38 [Pasteurella haemolytica]	56	38	870
511	1	1	1284	gi1217049	[brnQ protein [Salmonella typhimurium]	56	37	1284
604	3	1099	1701	gi1467109	[rimJ 10S ribosomal protein S18 alanine acetyltransferase; 229_C1_170 [Mycobacterium leprae]	56	43	603
660	5	3547	3774	gi1229106	[ZK930.1 [Caenorhabditis elegans]	56	30	228
707	1	35	400	gi1353929	[NADPH-sulfite reductase [flavoprotein component [Salmonella typhimurium]	56	38	366
709	2	1385	1095	gi13510801	[hydrogenase accessory protein [Methanococcus jannaschii]	56	38	291
718	1	1	495	gi1413948	[ipa-24d gene product [Bacillus subtilis]	56	35	495
744	1	87	677	gi1928836	[repressor protein [Lactococcus lactis phage BK5-T]	56	35	591
790	1	776	399	gi11511513	[ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	56	33	378
795	1	3	407	gi1205382	[cell division protein [Haemophilus influenzae]	56	34	405
813	1	19	930	gi1222161	[permease [Haemophilus influenzae]	56	28	912
855	1	3	515	gi11256621	[26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	56	33	513
968	1	2	466	gi1547513	[orf3 [Haemophilus influenzae]	56	37	465
973	2	1049	732	gi1886022	[MexR [Pseudomonas aeruginosa]	56	21	318
1203	1	5	223	gi1184251	[HHG-1 [Homo sapiens]	56	34	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1976	1	452	237	gi 9806	lysine-rich aspartic acid-rich protein [Plasmodium chabaudi] r[S22183]S22183 lysine/aspartic acid-rich protein - Plasmodium baudii	56	33	216
2161	1	2	400	gi 1237015	ORF4 [Bacillus subtilis]	56	27	399
2958	1	362	183	gi 466685	No definition line found [Escherichia coli]	56	26	180
2979	1	421	212	gi 1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	56	40	210
2994	2	526	326	gi 836646	phosphoribosylformimino-praia ketoisomerase [Rhodobacter phaeoides]	56	29	201
3026	1	179	328	gi 143306	penicillin V amidase [Bacillus sphaericus]	56	30	150
3149	1	289	146	gi 1166604	Similar to aldehyde dehydrogenase [Caenorhabditis elegans]	56	37	144
3770	1	63	401	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	56	43	339
4034	2	720	361	gi 1205355	[Na+/H+ antiporter [Haemophilus influenzae]	56	31	360
4145	1	1	324	gi 726095	long-chain acyl-CoA dehydrogenase [Mus musculus]	56	36	324
4200	1	505	254	gi 155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir[A42289]A42289 glucose-fructose oxidoreductase [EC 1.1.-.-] recursor - Zymomonas mobilis	56	40	252
4273	1	675	355	gi 308861	GTC start codon [Lactococcus lactis]	56	33	321
1	3	4095	3436	gi 5341	Putative orf YCLX8c, len:192 [Saccharomyces cerevisiae] r[S53591]S53591 hypothetical protein - yeast [Saccharomyces evisiae]	55	25	660
11	12	9377	8505	gi 216773	haloacetate dehalogenase H-1 [Halaxella sp.]	55	32	873
12	4	5133	4534	gi 467337	unknown [Bacillus subtilis]	55	26	600
19	5	5404	5844	gi 1001719	hypothetical protein [Synechocystis sp.]	55	25	441
23	13	14087	12339	gi 474190	lucA gene product [Escherichia coli]	55	30	1749
32	7	5368	6888	gi 1340096	unknown [Mycobacterium tuberculosis]	55	37	1521
34	3	2569	1808	gi 1303968	YqjQ [Bacillus subtilis]	55	39	762
34	5	3960	3412	gi 1303962	YqjK [Bacillus subtilis]	55	33	549
36	1	1291	647	gi 1606045	ORF_0118 [Escherichia coli]	55	27	645
36	6	6220	5243	gi 1001341	hypothetical protein [Synechocystis sp.]	55	31	978
47	3	3054	3821	gi 1001819	hypothetical protein [Synechocystis sp.]	55	21	768
49	1	2065	1127	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir[S37251]S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	55	36	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
67	11	8966	9565	gi1153053	[norA1199 protein (Staphylococcus aureus)]	55	23	600
75	3	881	1273	gi111698	[u-histidinol: NAD ⁺ oxidoreductase (EC 1.1.1.23) (aa 1-434) scherichia coli]	55	33	393
82	9	15387	14194	gi11136221	[carboxypeptidase (Sulfolobus solfataricus)]	55	35	1194
87	4	3517	4917	gi11064812	[function unknown (Bacillus subtilis)]	55	26	1401
88	2	1172	1636	gi1882463	[protein-Nip1-phosphohistidine-sugar phosphotransferase (Escherichia coli)]	55	35	465
92	1	127	516	gi1137832	[unknown (Bacillus subtilis)]	55	36	390
100	2	836	2035	gi1170274	[seaxanthin epoxidase (Nicotiana glauca)]	55	36	1200
100	5	5137	4658	gi1396660	[unknown open reading frame (Buchnera aphidicola)]	55	29	480
108	3	4266	2986	gi1149866	[M. jannaschii predicted coding region MJ1024 (Methanococcus jannaschii)]	55	31	1281
114	3	2616	1834	gi1151167	[formate dehydrogenase, alpha subunit (Methanococcus jannaschii)]	55	29	783
144	3	1805	1476	gi1100787	[unknown (Saccharomyces cerevisiae)]	55	35	330
165	5	6212	5508	gi11045884	[M. genitalium predicted coding region MG199 (Mycoplasma genitalium)]	55	27	705
189	5	2205	2576	gi1142569	[ATP synthase a subunit (Bacillus firmus)]	55	35	372
191	6	9136	6857	gi1559411	[80272.3 (Caenorhabditis elegans)]	55	39	2280
194	2	364	636	gi1145768	[K7 kinesin-like protein (Dictyostelium discoideum)]	55	34	273
209	4	1335	1676	gi1473357	[thi4 gene product (Schizosaccharomyces pombe)]	55	35	342
211	2	1693	1145	gi1410130	[ORF86 (Bacillus subtilis)]	55	37	549
213	2	644	1372	gi1633692	[TraA (Yersinia enterocolitica)]	55	28	729
214	7	4144	5481	gi11001793	[hypothetical protein (Synechocystis sp.)]	55	30	1338
221	7	11473	9197	gi1466520	[pocR (Salmonella typhimurium)]	55	32	2277
233	8	5908	4817	gi11237063	[unknown (Mycobacterium tuberculosis)]	55	38	1092
236	4	1375	2340	gi1146199	[putative (Bacillus subtilis)]	55	32	966
243	2	380	1885	gi1459907	[mercuric reductase (Plasmodium falciparum)]	55	29	1506
258	1	786	394	gi1455006	[orf6 (Rhodococcus fascians)]	55	36	393
281	1	126	938	gi11408493	[homologous to Swisaprot:YIDA_ECOLI hypothetical protein (Bacillus subtilis)]	55	35	813
316	3	1323	2102	gi11486447	[LuxA homologue (Rhizobium sp.)]	55	30	780
326	5	2968	2744	gi11296824	[proline iminopeptidase (Lactobacillus helveticus)]	55	36	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
351	2	2322	1429	gi 1204820	hydrogen peroxide-inducible activator [Haemophilus influenzae]	55	28	894
353	4	2197	2412	gi 1272475	chitin synthase [Emericella nidulans]	55	50	216
380	1	14	379	gi 142554	ATP synthase 1 subunit [Bacillus megaterium]	55	37	366
383	1	462	232	gi 198972	ferrichrome-binding protein [Bacillus subtilis]	55	36	231
386	1	3	938	gi 1510251	DNA helicase, putative [Methanococcus jannaschii]	55	30	936
410	2	1208	1891	gi 1205144	multidrug resistance protein [Haemophilus influenzae]	55	27	684
483	2	411	833	gi 113934	ipa-10r gene product [Bacillus subtilis]	55	26	423
529	3	1777	1433	gi 606150	ORF_1309 [Escherichia coli]	55	33	345
555	1	1088	585	gi 143407	para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]	55	28	504
565	1	402	202	gi 1223961	CDP-tylucose epimerase [Yersinia pseudotuberculosis]	55	41	201
582	1	751	852	gi 1256643	20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion; putative [Bacillus subtilis]	55	36	300
645	5	2260	2057	gi 210824	fusion protein F (Bovine respiratory syncytial virus) p1r JQ1481 VONZBA fusion glycoprotein precursor - bovine respiratory syncytial virus (strain A51908)	55	25	204
672	2	957	2216	gi 1511333	M. jannaschii predicted coding region M1122 [Methanococcus jannaschii]	55	36	1260
730	1	955	479	gi 537007	ORF_1379 [Escherichia coli]	55	30	477
737	1	1859	945	gi 536963	CG Site No. 18166 [Escherichia coli]	55	30	915
742	2	228	572	gi 304160	product unknown [Bacillus subtilis]	55	38	345
817	2	1211	903	gi 1136289	histidine kinase A [Dictyostelium discoideum]	55	29	309
819	1	582	355	gi 558073	polymorphic antigen [Plasmodium falciparum]	55	22	228
832	2	1152	724	gi 40367	ORF [Clostridium acetobutylicum]	55	32	429
840	1	769	386	gi 1205875	pseudouridylylase synthase 1 [Haemophilus influenzae]	55	39	384
1021	1	23	529	gi 48563	beta-lactamase [Yersinia enterocolitica]	55	38	507
1026	1	60	335	gi 47804	Opp C (AAI-301) [Salmonella typhimurium]	55	26	276
1525	1	1	282	gi 1477533	IsaA [Staphylococcus aureus]	55	29	282
1814	2	224	985	gi 1046078	M. genitalium predicted coding region MG369 [Mycoplasma genitalium]	55	38	762
3254	1	427	254	gi 413968	ipa-44d gene product [Bacillus subtilis]	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3695	1	686	345	gi 216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	342
3721	1	1	312	gi 42029	ORF1 gene product [Escherichia coli]	55	31	312
3799	1	3	272	gi 42029	ORF1 gene product [Escherichia coli]	55	38	270
3889	1	22	423	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	45	402
3916	1	2	385	gi 529754	speC (Streptococcus pyogenes)	55	38	384
3945	1	4	198	gi 476252	phase 1 flagellin [Salmonella enterica]	55	36	195
4074	1	488	246	gi 42029	ORF1 gene product [Escherichia coli]	55	38	243
4184	1	2	343	gi 1524267	unknown Mycobacterium tuberculosis	55	28	342
4284	1	14	208	gi 1100774	ferredoxin-dependent glutamate synthase [Synecocystis sp.]	55	36	195
4457	2	644	378	gi 180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi 182737 cerebellar degeneration-associated protein [Homo sapiens] p1r A29770 A29770 cerebellar degeneration-related protein - human	55	38	267
4514	1	2	244	gi 216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	243
4599	1	432	217	gi 1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	42	216
4606	1	416	210	gi 386120	myosin alpha heavy chain (52 subfragment) [rabbits, masseter, eptide Partial, 234 aa]	55	27	207
5	8	5348	4932	gi 536069	ORF YBL047c [Saccharomyces cerevisiae]	54	27	417
12	7	7166	6165	gi 1205504	homoserine acetyltransferase [Haemophilus influenzae]	54	30	1002
23	16	17086	15326	gi 474192	lucC gene product [Escherichia coli]	54	31	1761
35	1	2	979	gi 48054	small subunit of soluble hydrogenase (AA 1-384) [Synecococcus sp.] lr S06919 HQYCS soluble hydrogenase (EC 1.12.-.-) small chain - unclonococcus sp. (PCC 6716)	54	36	978
37	11	9437	8667	gi 537207	ORF_f277 [Escherichia coli]	54	38	771
37	12	8165	8332	gi 1160967	palmitoyl-protein thioesterase [Homo sapiens]	54	37	168
46	15	13025	13804	gi 438473	protein is hydrophobic, with homology to E. coli ProW; putative Bacillus subtilis	54	28	780
56	2	203	736	gi 1256139	ybbJ [Bacillus subtilis]	54	34	534
57	13	11117	10179	gi 1151248	inosine-uridine preferring nucleoside hydrolase [Crithidia fasciculata]	54	32	939
66	2	516	1133	gi 1335781	Cap [Drosophila melanogaster]	54	29	618
70	10	8116	8646	gi 1399823	PhoE [Rhizobium meliloti]	54	31	531

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
70	15	12556	11601	[sp P02983 TCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	29	756
87	5	4915	5706	[gi 1064811	[function unknown [Bacillus subtilis]	54	33	792
92	4	3005	2289	[gi 1205366	[oligopeptide transport ATP-binding protein [Haemophilus influenzae]	54	33	717
103	2	2596	1556	[gi 710495	[protein kinase [Bacillus brevis]	54	33	1041
105	2	3585	2095	[gi 143727	[putative [Bacillus subtilis]	54	30	1491
112	4	2337	2732	[gi 153724	[MalC [Streptococcus pneumoniae]	54	41	396
127	2	1720	2493	[gi 144297	[acetyl esterase (XyxC) [Caldocellum saccharolyticum] pir B37203 B37202 [acetyl esterase (EC 3.1.1.6) (XyxC) - Caldocellum saccharolyticum]	54	34	774
138	5	1600	3306	[gi 42473	[pyruvate oxidase [Escherichia coli]	54	36	1707
152	2	525	1172	[gi 1377834	[unknown [Bacillus subtilis]	54	23	648
161	9	4831	5469	[gi 903305	[ORF73 [Bacillus subtilis]	54	28	639
161	13	6694	7251	[gi 1511039	[phosphate transport system regulatory protein [Methanococcus jannaschii]	54	32	558
164	6	3263	4543	[gi 1204976	[prolyl-tRNA synthetase [Haemophilus influenzae]	54	34	1281
164	20	21602	22243	[gi 143582	[apollitea protein [Bacillus subtilis]	54	32	642
171	6	5683	4250	[gi 436965	[malA] gene products [Bacillus stearothermophilus] pir S43914 S43914 [hypothetical protein 1 - Bacillus stearothermophilus]	54	37	1434
206	18	19208	19720	[gi 1240016	[R09E10.3 [Caenorhabditis elegans]	54	38	533
218	2	1090	1505	[gi 467378	[unknown [Bacillus subtilis]	54	26	816
220	1	1322	663	[gi 1353761	[myosin II heavy chain [Naegleria fowleri]	54	22	660
220	13	12655	13059	[pir S00485 S004	[gene 11-1 protein precursor - Plasmodium falciparum (fragments)]	54	35	405
221	3	2030	3709	[gi 1303813	[Xqew [Bacillus subtilis]	54	34	1680
272	7	5055	4219	[gi 62964	[arylamine N-acetyltransferase (AA 1-290) [Gallus gallus] ir S06652 XVCHY3 [arylamine N-acetyltransferase (EC 2.3.1.5) (clone NAT-3) - chicken]	54	33	837
316	7	4141	4701	[gi 682769	[accE gene product [Escherichia coli]	54	31	561
316	10	6994	8742	[gi 413931	[ipa-27d gene product [Bacillus subtilis]	54	28	1749
338	3	3377	2214	[gi 490328	[LORF F [unidentified]	54	28	1164
341	4	3201	3614	[gi 171959	[myosin-like protein [Saccharomyces cerevisiae]	54	25	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
346	1	1820	912	gi 396400	similar to eukaryotic Na ⁺ /H ⁺ exchangers [Escherichia coli] sp P22703 XCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOKR-ACS MYRGENIC REGION (0549).	54	34	909
348	2	623	1351	gi 537109	ORF_f343a [Escherichia coli]	54	34	729
378	2	1007	1942	sp P02983 TCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	31	936
408	6	4351	5301	gi 474190	lucA gene product [Escherichia coli]	54	29	951
444	9	7934	8854	gi 216267	ORF2 [Bacillus megaterium]	54	32	921
463	2	2717	2229	gi 104160	product unknown [Bacillus subtilis]	54	50	489
502	2	1696	1133	gi 1205015	hypothetical protein (sp P10120) [Haemophilus influenzae]	54	38	564
505	6	6262	5357	gi 1500558	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase [Methanococcus jannaschii]	54	41	906
550	1	2736	1522	gi 40100	rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] Ir S06049 S06049 rodC protein - Bacillus subtilis p P13485 TAGP_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.	54	35	1215
551	5	3305	4279	gi 950197	unknown [Corynebacterium glutamicum]	54	34	975
558	2	1356	958	gi 485090	No definition line found [Caenorhabditis elegans]	54	32	399
580	1	91	936	gi 311906	fused envelope glycoprotein precursor [Friend spleen focus-forming virus]	54	45	846
603	3	554	757	gi 1323423	ORF YGR334w [Saccharomyces cerevisiae]	54	36	204
617	1	25	249	gi 219959	ornithine transcarbamylase [Homo sapiens]	54	40	225
622	3	1097	1480	gi 1303873	Ygg2 [Bacillus subtilis]	54	25	384
623	1	3	404	gi 1063250	low homology to p20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillus [Bacillus subtilis]	54	45	402
689	1	1547	1011	gi 552446	NADH dehydrogenase subunit 4 [Apis mellifera ligustical] pir S52968 S52968 NADH dehydrogenase chain 4 - honeybee itochondrion (SC4)	54	30	537
725	2	686	1441	gi 987096	sensory protein kinase [Streptomyces hygroscopicus]	54	26	756
956	1	1	249	pir S30782 S307	Integrin homolog - yeast [Saccharomyces cerevisiae]	54	24	249
978	2	1137	859	gi 1301994	ORF YNL091w [Saccharomyces cerevisiae]	54	33	279
1314	1	3	281	gi 1001108	hypothetical protein [Synecocystis sp.]	54	33	279
2450	1	1	228	gi 1045057	Ch-TOG [Homo sapiens]	54	32	228
2934	1	1	387	gi 580870	ipe-37d qoxA gene produ... [Bacillus subtilis]	54	36	387
2970	1	499	251	sp P3734P VECE_	HYPOTHETICAL PROTEIN IN ASP5 5'-REGION (FRAGMENT).	54	42	249

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3002	1	1	309	gi 44027	Tma protein [Lactococcus lactis]	54	33	309
3561	1	9	464	gi 151359	HMO-CoA reductase (EC 1.1.1.88) [Pseudomonas nevalonii] pir A44756 A44756 hydroxyethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	54	35	456
3572	1	72	401	gi 450688	hcdm gene of Ecopri gene product [Escherichia coli] pir S38437 S38437 hadh protein - Escherichia coli; pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	54	36	330
3829	1	798	400	gi 1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	54	29	399
3909	1	1	273	gi 29865	CENP-E (Homo sapiens)	54	30	273
3921	1	3	209	pir S24325 S243	glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa	54	34	207
4438	1	566	285	gi 1196657	unknown protein [Mycoplasma pneumoniae]	54	30	282
4459	1	3	272	gi 1046081	hypothetical protein (G8:D26185_10) [Mycoplasma genitalium]	54	38	270
4564	1	3	221	gi 216267	ORP2 [Bacillus megaterium]	54	38	219
23	12	12538	10685	gi 474192	lucC gene product [Escherichia coli]	53	35	1854
23	14	14841	13579	gi 46029	ORP1 gene product [Escherichia coli]	53	32	1263
24	3	4440	3940	gi 1369947	lc2 gene product [Bacteriophage B1]	53	36	501
26	4	3818	4618	gi 1486247	unknown [Bacillus subtilis]	53	37	801
38	6	2856	3998	gi 405880	lyeII [Escherichia coli]	53	40	1143
38	10	9380	7806	gi 1399954	thyroid sodium/iodide symporter NIS [Rattus norvegicus]	53	29	1575
56	10	12324	12100	pir A54592 A545	110k actin filam... associated protein - chicken	53	32	225
57	6	5047	4583	pir A00341 DEZP	alcohol dehydrogenase (EC 1.1.1.1) - fission yeast (Schizosaccharomyces pombe)	53	39	465
57	12	110515	8932	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	30	1584
67	12	9496	10218	gi 1511555	quinolone resistance nraA protein protein [Methanococcus jannaschii]	53	31	723
69	3	3125	2382	gi 1687017	arabinogalactan-protein, AGP [Nicotiana glauca, cell-suspension culture filtrate, Peptide, 461 aa]	53	30	744
79	1	3	1031	gi 1523802	glucanase [Anabaena variabilis]	53	32	1029
80	1	673	338	gi 452428	ATPase 3 [Plasmodium falciparum]	53	36	336
88	4	1910	2524	gi 137034	ORP_0488 [Escherichia coli]	53	25	615
88	5	2467	3282	gi 1537034	ORP_0488 [Escherichia coli]	53	29	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match	gene name	% sim	% ident	length (nt)
92	8	5870	5505	gi 399598	amphotropic murine retrovirus receptor [Rattus norvegicus]	53	33	366
94	5	4417	3239	gi 173038	tropomyosin (TPM) [Saccharomyces cerevisiae]	53	25	1179
99	5	4207	5433	sp P28246 BCR_E	BICYCLOMYCIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN).	53	30	1227
120	3	1639	2262	gi 576655	ORF1 [Vibrio anguillarum]	53	35	624
120	11	7257	8897	gi 1524397	glycine betaine transporter Opud [Bacillus subtilis]	53	33	1641
127	6	6893	5685	gi 1256630	putative [Bacillus subtilis]	53	32	1209
147	2	255	557	gi 581648	epib gene product [Staphylococcus epidermidis]	53	34	303
158	4	4705	4256	gi 151004	mucoidy regulatory protein AlgR [Pseudomonas aeruginosa] pir A32802 A32802 regulatory protein algR - Pseudomonas aeruginosa sp P26275 ALGR_PSEAE POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN.	53	32	450
171	7	5717	5421	gi 1510669	hypothetical protein (GP-D6404_18) [Methanococcus jannaschii]	53	34	297
191	9	113087	11483	gi 298085	acetate decarboxylase [Clostridium acetobutylicum] pir B49346 B49346 butyrate-acetate CoA-transferase (EC 2.8.3.9) small chain - Clostridium acetobutylicum sp P33752 CTFA_CLOAB BUTYRATE-ACETOACETATE COA-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)	53	31	1605
203	5	3763	4326	gi 143456	lipoE protein (cgt start codon) [Bacillus subtilis]	53	29	564
206	17	118204	118971	gi 304136	acetylglutamate kinase [Bacillus stearothermophilus] sp D07905 ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (AGK) (N-ACETYL-L-GLUTAMATE 5-PHOSPHOTRANSFERASE).	53	36	768
212	10	4021	4221	gi 9878	protein kinase [Plasmodium falciparum]	53	28	201
231	2	1580	1350	gi 537506	paramyosin [Dirofilaria immitis]	53	34	231
272	6	2719	3249	pir A33141 A331	hypothetical protein (cgtf 3' region) - Streptococcus mutans	53	34	531
308	3	927	2576	gi 606292	ORF_0696 [Escherichia coli]	53	33	1630
320	7	5645	5884	gi 160596	RNA polymerase III largest subunit [Plasmodium falciparum] sp P27625 RPCL_PLAPA DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6).	53	33	240
327	1	218	901	gi 854601	unknown [Schizosaccharomyces pombe]	53	31	684
341	2	212	2500	gi 633732	ORF1 [Campylobacter jejuni]	53	31	2289
351	1	763	383	sp P31675 YABM_	HYPOTHEICAL 42.7 KD PROTEIN IN TBPA-LEUD INTERGENIC REGION (ORF104).	53	32	381
433	7	5087	4731	gi 1001561	MHC class II analog [Staphylococcus aureus]	53	30	357
454	2	1240	980	pir A60328 A603	40K cell wall protein precursor (at 5' region) - Streptococcus mutans (strain OH2175, serotype f)	53	27	261

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	RefSeq gene name	% sim	% ident	length (nt)
470	4	1123	1761	gi1516826	rat GCP360 (Rattus rattus)	53	30	639
483	1	432	217	gi11480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	31	216
544	1	516	1259	gi146587	ORF 1 (AA 1 - 121) (1 is 2nd base in codon) [Staphylococcus aureus] if [S15765]S15765 hypothetical protein 1 (hnb 5' region) - aphylococcus aureus (fragment)	53	38	744
558	10	3957	3754	gi115140	tes gene [Bacteriophage P1]	53	32	204
603	2	339	620	gi1507738	Hmp [Vibrio parahaemolyticus]	53	26	282
693	1	1669	941	gi1153123	toxic shock syndrome toxin-1 precursor [Staphylococcus aureus] p1r[A24606]XCSA31 toxic shock syndrome toxin-1 precursor - Laphylococcus aureus	53	38	729
766	1	2	673	gi1687600	orfA2; orfA2 forms an operon with orfA1 [Listeria monocytogenes]	53	43	672
781	1	667	335	gi11204551	pilin biogenesis protein [Haemophilus influenzae]	53	26	333
801	1	3	545	gi11279400	SapA protein [Escherichia coli]	53	25	543
803	1	2	910	gi1695278	lipase-like enzyme [Alcaligenes eutrophus]	53	30	909
872	1	1177	590	gi1298032	EP [Streptococcus suis]	53	30	588
910	1	2	184	gi11044936	unknown [Schistosoma caracasense]	53	29	183
943	1	794	399	gi1390508	similar to unidentified ORF near 47 minutes [Escherichia coli] sp[P31436]VICK_ECOLI HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA NTERGENIC REGION.	53	30	396
988	1	1004	504	gi1142441	ORF 3; putative [Bacillus subtilis]	53	28	501
1064	1	3	434	gi1305080	myosin heavy chain [Entamoeba histolytica]	53	26	432
1366	1	3	452	gi1308852	transmembrane protein [Lactococcus lactis]	53	33	450
1758	1	792	397	gi11001774	hypothetical protein [Synecocystis sp.]	53	30	396
1897	1	1	447	gi11303949	Yqix [Bacillus subtilis]	53	27	447
2381	1	798	400	gi11146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	53	37	399
3537	1	1	327	gi1450688	hsm gene of Ecoor1 gene product [Escherichia coli] p1r[S38437]S38437 hdm protein - Escherichia coli p1r[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-320)	53	35	327
3747	2	137	397	gi11477486	transposase [Burkholderia cepacia]	53	53	261
11	5	3049	3441	gi1868224	No definition line found [Caenorhabditis elegans]	52	33	393

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
15	5	2205	2369	gi 215966	G41 protein (gag start codon) [Bacteriophage T4]	52	34	165
19	3	2429	3808	gi 1205379	UDP-murac-pentapeptide synthetase [Haemophilus influenzae]	52	31	1380
24	1	6920	3462	gi 579124	predicted 86.4kd protein: 52kd observed [Mycobacteriophage L5] p1r[530971 530971] gene 26 protein - Mycobacterium phage L5 sp[Q05233 V026_BPHUS MINOR TAIL PROTEIN QP26. (SUB 2-837)	52	32	3459
37	5	3015	3935	gi 1500543	p115 protein [Methanococcus jannaschii]	52	25	921
38	13	8795	9703	gi 46851	glucose kinase [Streptomyces coelicolor]	52	29	909
44	16	10617	11066	gi 42012	moaE gene product [Escherichia coli]	52	36	450
46	1	3	521	gi 1040957	NADH dehydrogenase subunit 6 [Anopheles trinkae]	52	25	519
51	10	5531	6280	gi 388269	lcrC [Plasmodium]	52	32	750
56	5	3968	2826	gi 181949	endothelial differentiation protein (edg-1) [Homo sapiens] p1r[A35300 A35300 G protein-coupled receptor edg-1 - human sp[P21453 EDG1_HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi 304153	sorbitol dehydrogenase [Bacillus subtilis]	52	27	678
62	5	3364	2870	gi 1072399	phaE gene product [Rhizobium meliloti]	52	25	495
62	6	4445	3651	gi 46485	NADH dehydrogenase [Synecococcus PCC7942]	52	27	795
67	14	11355	12962	gi 1511365	glutamate synthase (NADPH), subunit alpha [Methanococcus jannaschii]	52	30	1608
67	21	16935	18158	gi 1204393	hypothetical protein (SP:P31122) [Haemophilus influenzae]	52	25	1224
70	4	2185	1997	gi 7227	cytoplasmic dynein heavy chain [Dictyostelium discoideum] r[A44357 A44357 cytoplasmic dynein heavy chain, cytosolic - slime mold cytosolium discoideum]	52	36	189
96	10	10005	10664	gi 1408485	1865G gene product [Bacillus subtilis]	52	26	660
103	5	3986	3351	gi 1009368	Respiratory nitrate reductase [Bacillus subtilis]	52	42	636
109	3	4102	3350	gi 699274	lmbE gene product [Mycobacterium leprae]	52	39	753
109	19	15732	17300	gi 1526981	amino acid permease YeeF like protein [Salmonella typhimurium]	52	30	1569
121	3	1412	981	gi 732931	unknown [Saccharomyces cerevisiae]	52	32	432
125	3	865	1680	gi 1296975	puf gene product [Porphyromonas gingivalis]	52	38	816
130	2	659	1807	gi 1256624	25.8% identity over 120 aa with the Synecococcus sp. MpeV protein; putative [Bacillus subtilis]	52	36	1149
149	1	1164	583	gi 1225943	PBSX terminase [Bacillus subtilis]	52	33	582
149	16	4687	4415	gi 1510368	M. jannaschii predicted coding region MJO272 [Methanococcus jannaschii]	52	35	273

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
167	1	216	1001	gi 146025	cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi 474915	orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin resistance protein of Escherichia coli [Coxiella burnetii] pir[S44207/S44207] hypothetical protein 337 - Coxiella burnetii (SUB -338)	52	26	1137
195	9	9161	8760	gi 3028	mitochondrial outer membrane 72K protein [Neurospora crassa] r[AJ6682/AJ6682 72K mitochondrial outer membrane protein - neurospora crassa]	52	25	402
200	3	2065	2607	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	543
203	4	2776	3684	gi 1303698	bltD [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi 305080	myosin heavy chain [Entamoeba histolytica]	52	24	402
242	1	21	1424	gi 1060877	[EmrY [Escherichia coli]	52	32	1404
249	5	4526	4753	pir[C37222/C372	cytochrome P450 1A1, hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi 143290	penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi 1001610	hypothetical protein [Synchocystis sp.]	52	30	300
276	8	4456	4055	gi 416235	orf L3 [Mycoplasma capricolum]	52	26	402
289	2	1856	1449	gi 150900	GTP phosphohydrolase [Proteus vulgaris]	52	34	408
325	1	1	279	gi 1204874	polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi 1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	52	33	1008
375	3	340	1878	gi 467446	similar to SpvB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi 1478239	unknown [Mycobacterium tuberculosis]	52	34	843
430	1	3	575	pir[A42606/A426	orfA 5' - orfA05 - Saccharopolyspora erythraea (fragment)	52	28	573
444	4	4728	3712	gi 1408494	homologous to penicillin acylase [Bacillus subtilis]	52	31	1017
465	1	1802	903	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir[A27650/A27650 regulatory protein phoR - Bacillus subtilis sp P23545 PHO_R_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-)]	52	36	900
469	5	4705	4169	gi 755152	highly hydrophobic integral membrane protein [Bacillus subtilis] sp P42953 TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	52	32	537
495	1	1262	631	gi 1204607	transcription activator [Haemophilus influenzae]	52	25	630
505	7	6004	5762	gi 142440	ATP-dependent nuclease [Bacillus subtilis]	52	28	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
517	2	1162	1614	gi1166162	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	52	35	453
543	2	444	1295	gi1215693	putative orf; G9_0r1434 [Mycoplasma pneumoniae]	52	25	852
586	1	1	336	gi1581648	epib gene product [Staphylococcus epidermidis]	52	36	336
773	1	848	426	gi1279769	YdhC [Methanobacterium thermoformicum]	52	30	423
1120	2	100	330	gi142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	231
1614	1	691	347	gi1289262	comE ORF [Bacillus subtilis]	52	28	345
2495	1	1	324	gi1216151	DNA polymerase (gene L; tlg start codon) [Bacteriophage SP02] gi1579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pirA21498 [DJBP52 DNA-directed DNA polymerase (EC 2.7.7) - phage P02]	52	34	324
2931	1	566	285	gi1256136	YbbG [Bacillus subtilis]	52	30	282
2943	1	577	320	gi141713	hlsA ORF (AA 1-245) [Escherichia coli]	52	35	258
2993	1	588	295	gi1298032	EF [Streptococcus suis]	52	34	294
3667	1	612	307	gi1849025	hypothetical 64.7-KDa protein [Bacillus subtilis]	52	36	306
3944	1	478	260	gi1218040	BAA [Bacillus licheniformis]	52	36	219
3954	2	613	347	gi1854064	[U87 Human herpesvirus 6]	52	50	267
3986	1	90	403	gi1205919	Na ⁺ and Cl ⁻ dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	52	33	312
4002	1	3	389	gi140003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p123129 [ODOI_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	52	42	387
4020	1	1	249	gi159388	ornithine decarboxylase [Leishmania donovani]	52	47	249
4098	1	438	220	gi1409795	No definition line found [Escherichia coli]	52	32	219
4248	1	3	212	gi1965077	AdpA [Saccharomyces cerevisiae]	52	40	210
7	1	3	575	gi1895747	putative cel operon regulator [Bacillus subtilis]	51	28	573
21	4	2479	3276	gi1510962	indole-3-glycerol phosphate synthase [Methanococcus jannaschii]	51	32	798
22	9	5301	5966	gi1303933	YqjN [Bacillus subtilis]	51	25	666
43	3	1516	1283	gi1519460	Srp1 [Schizosaccharomyces pombe]	51	31	234
44	17	11042	11305	gi142011	moaD gene product [Escherichia coli]	51	35	264
51	11	6453	6731	gi1495471	vacuolating toxin [Helicobacter pylori]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
52	4	2537	2995	gi 1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	51	32	459
57	10	7331	6843	gi 508173	SIIA domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	51	32	489
59	1	29	1111	gi 299163	alanine dehydrogenase [Bacillus subtilis]	51	33	1083
67	20	15791	16576	gi 1510977	M. jannaschii predicted coding region MJ0938 [Methanococcus jannaschii]	51	24	786
69	2	1559	1218	gi 467359	unknown [Bacillus subtilis]	51	34	342
71	1	3	1196	gi 298032	BP [Streptococcus suis]	51	32	1194
78	2	349	176	gi 1161242	proliferating cell nuclear antigen [Styela clava]	51	28	174
99	4	3357	4040	gi 642795	TFIID subunit TAF155 [Homo sapiens]	51	25	684
109	1	2852	1428	gi 580920	rodd (GTA) polypeptide (AA 1-673) [Bacillus subtilis] pir[S06048]S06048 probable rodd protein - bacillus subtilis sp P13484 TAGE.BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHN-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTIN E).	51	27	1425
109	9	6007	6693	gi 1204815	hypothetical protein (SP:P32662) [Haemophilus influenzae]	51	23	687
112	3	1066	2352	pir[S05310]S053	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	14432	12855	gi 405857	yehU [Escherichia coli]	51	29	1578
114	9	9725	8967	gi 435098	lorf1 [Mycoplasma capricolum]	51	30	759
115	1	1	912	gi 1431110	ORF YH085w [Saccharomyces cerevisiae]	51	25	912
127	10	9647	10477	gi 1204314	H. influenzae predicted coding region H10056 [Haemophilus influenzae]	51	37	831
152	9	6814	7356	gi 431929	Munf regulatory protein [Mycoplasma sp.]	51	38	543
154	2	575	1153	gi 1237044	unknown [Mycobacterium tuberculosis]	51	36	579
154	7	6587	5634	gi 409286	batU [Bacillus subtilis]	51	27	954
171	8	6943	6236	gi 1205484	hypothetical protein (SP:P33918) [Haemophilus influenzae]	51	32	708
184	1	1	291	gi 466886	bl496_C3_206 [Mycobacterium leprae]	51	33	291
212	5	1501	2139	pir A45605 A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi 8204	nuclear protein [Drosophila melanogaster]	51	27	672
236	8	8137	7481	gi 49272	Asparaginase [Bacillus licheniformis]	51	31	657
243	4	4637	3546	gi 1511102	malvalonate kinase [Methanococcus jannaschii]	51	29	1092

TABLE 2

TABLE 2

Contig ID	ORF ID	Start (nt)	Stop (nt)	match %	match position	blast gene name	% sim	% ident	length (nt)
257	4	3540	3373	gi 204579		H. Influenzae predicted coding region H1026 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	gi 160299		glutamic acid-rich protein [Plasmodium falciparum] pir [A54514][A54514] glutamic acid-rich protein precursor - Plasmodium alciparum	51	34	789
255	5	2419	3591	gi 580841		Pf [Bacillus subtilis]	51	32	1173
298	2	518	748	gi 1336182		SCP8 [Streptococcus agalactiae]	51	34	231
316	9	5817	7049	gi 413953		Ipa-29D gene product [Bacillus subtilis]	51	39	1233
332	2	3775	2057	gi 1209012		[mutS] [Thermus aquaticus thermophilus]	51	26	1719
364	4	3816	4991	gi 528991		unknown [Bacillus subtilis]	51	32	1176
440	2	448	684	gi 7819		[transferase (GAL10) (AA 1 - 687)] [Kluyveromyces lactis] r[S01407]XUVKG UDPglucose 4-epimerase (EC 5.1.1.2) - yeast uyveromyces maritimus var. lactis]	51	32	237
495	2	1353	1177	gi 297861		[protease G [Erwinia chrysanthemi]]	51	41	177
495	3	2287	1718	gi 1513317		[serine rich protein [Entamoeba histolytica]]	51	25	570
506	1	840	421	gi 455320		[cII protein [Bacteriophage P4]]	51	33	420
600	1	1474	983	gi 587532		[orf, len: 201, CAX: 0.16 [Saccharomyces cerevisiae] pir [S48818][S48818] hypothetical protein - yeast (Saccharomyces erevisiae)]	51	30	492
607	3	479	934	gi 1511524		[hypothetical protein (SP:P37002) [Methanococcus jannaschii]]	51	40	456
686	2	127	600	gi 493017		[endocarditis specific antigen [Enterococcus faecalis]]	51	30	474
726	1	33	230	gi 1353851		[unknown [Prochlorococcus marinus]]	51	45	198
861	1	176	652	gi 410145		[dehydroquinase dehydratase [Bacillus subtilis]]	51	34	477
869	1	782	393	gi 40100		[rodc (tag3) polypeptide (AA 1-746) [Bacillus subtilis] ic[S06049]S06049 rodc protein - Bacillus subtilis p[PIJ485]TAGP_BACSDU TECHNOIC ACID BIOSYNTHESIS PROTEIN F.	51	23	390
1001	1	642	322	gi 1279707		[hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]]	51	39	321
1046	2	866	624	gi 510257		[glycosyltransferase [Escherichia coli]]	51	29	243
1467	1	702	352	gi 1511175		[M. jannaschii predicted coding region MJ1177 [Methanococcus jannaschii]]	51	32	351
2558	1	457	230	sp P10582 DPOM_		[DNA POLYMERASE (EC 2.7.7.7) (S-1 DNA ORF 3).	51	26	228
3001	1	779	399	gi 1809543		[CbrC protein [Erwinia chrysanthemi]]	51	27	381
3604	1	1	399	pi JC4210 JC42		[3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse	51	37	399
3732	1	2	316	gi 145906		[acyl-CoA synthetase [Escherichia coli]]	51	33	315

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3791	1	2	274	gi 1061351	semaphorin III family homolog [Homo sapiens]	51	37	273
3995	1	46	336	gi 216346	surfactin synthetase [Bacillus subtilis]	51	38	291
4193	1	612	307	gi 42749	ribosomal protein L12 (AA 1-179) [Escherichia coli] r S04776 XXBEP2 peptide N-acetyltransferase rimL (EC 2.3.1.-) - cherichia coli	51	25	306
4539	1	367	185	gi 1408494	homologous to penicillin acylase [Bacillus subtilis]	51	40	183
4562	1	442	239	gi 1458280	coded for by C. elegans cDNA cm01e7; Similar to hydroxymethylglutaryl-CoA synthase [Caenorhabditis elegans]	51	35	204
1	4	3576	4859	gi 559160	GRAIL score: null; cap site and late promoter motifs present putative [Autographa californica nuclear polyhedrosis virus]	50	44	1284
11	7	4044	5165	gi 1146207	putative [Bacillus subtilis]	50	35	1122
11	13	10509	9496	gi 1208451	hypothetical protein [Synechocystis sp.]	50	39	1016
19	1	2034	1018	gi 413966	ipe-42d gene product [Bacillus subtilis]	50	29	1017
20	11	8586	8407	gi 1323159	ORF YGR103W [Saccharomyces cerevisiae]	50	28	180
24	5	5408	4824	gi 496280	structural protein [Bacteriophage Tuc2009]	50	29	585
34	4	1926	2759	gi 1303966	YqjO [Bacillus subtilis]	50	36	834
38	30	22865	23440	gi 1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	576
47	2	1705	2976	gi 153015	FemA protein [Staphylococcus aureus]	50	29	1272
56	13	15290	15841	gi 606096	ORF_f167; end overlaps end of o100 by 14 bases; start overlaps t174, ther starts possible [Escherichia coli]	50	30	562
57	1	2135	1077	gi 640922	xyitol dehydrogenase (unidentified hemiascomycete)	50	29	1059
58	2	628	1761	gi 143725	putative [Bacillus subtilis]	50	29	1134
88	6	4393	3884	gi 1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	510
89	5	3700	3356	gi 1276658	ORF174 gene product [Porphyra purpurea]	50	25	345
141	1	3	239	gi 476024	carbamoyl phosphate synthetase II [Plasmodium falciparum]	50	33	237
151	1	186	626	gi 1403441	unknown [Mycobacterium tuberculosis]	50	35	441
166	7	11065	9623	gi 895747	putative cel operon regulator [Bacillus subtilis]	50	32	1443
201	6	5284	5096	gi 160229	circumsporozoite protein [Plasmodium reichenowi]	50	42	189
206	12	30784	29555	gi 1052754	LarP integral membrane protein [Lactococcus lactis]	50	24	1230

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
211	4	1523	1927	gi 410131	ORF7 [Bacillus subtilis]	50	29	405
214	4	2411	3295	sp P37348 VECE_	HYPOTHETICAL PROTEIN IN ASP5 5'REGION (FRAGMENT).	50	37	885
228	7	5068	4406	gi 313580	envelope protein [Human immunodeficiency virus type 1] pif[S35835]S35835 envelope protein - human immunodeficiency virus type 1 (fragment) (SUB 1-77)	50	35	663
272	2	3048	1723	gi 1408485	B65G gene product [Bacillus subtilis]	50	22	1326
273	2	1616	984	gi 184186	phosphoglycerate mutase [Saccharomyces cerevisiae]	50	28	633
328	2	2507	1605	gi 148896	lipoprotein [Haemophilus influenzae]	50	26	903
332	4	5469	3802	gi 1526547	DNA polymerase family X [Thermus aquaticus]	50	27	1668
342	5	3473	3931	gi 456562	C-box binding factor [Dictyostelium discoideum]	50	35	459
352	1	1478	741	gi 288301	ORF2 gene product [Bacillus megaterium]	50	29	738
408	7	5299	5523	gi 11665	ORF2136 [Marchantia polymorpha]	50	27	225
420	3	650	1825	gi 757842	UDP-sugar hydrolase [Escherichia coli]	50	30	1176
464	1	1	591	gi 487282	Na ⁺ -ATPase subunit J [Enterococcus hirae]	50	29	591
472	2	1418	864	gi 551875	BglR [Lactococcus lactis]	50	23	555
520	1	23	541	gi 567036	Cape [Staphylococcus aureus]	50	27	519
529	1	6	410	gi 1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	50	34	405
534	5	7726	6059	gi 1255671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]	50	18	1668
647	1	2990	1497	gi 405568	TraI protein shares sequence similarity with a family of topoisomerases [Plasmid pSK41]	50	31	1494
664	3	1133	713	gi 410007	Leukocidin F component [Staphylococcus aureus, MRSA No. 4, Peptide, 23 aa]	50	32	423
678	1	1	627	gi 238032	EF [Streptococcus suis]	50	29	627
755	3	947	1171	gi 150572	cytochrome c1 precursor (EC 1.10.2.2) [Paracoccus denitrificans] gi 45465 cytochrome c1 (AA 1-450) [Paracoccus denitrificans] pif[C29413]C29413 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1 precursor - Paracoccus denitrificans sp 3627 CV1	50	37	225
827	1	1363	683	gi 142020	heterocyst differentiation protein [Anabaena sp.]	50	21	681
891	1	3	752	gi 1408485	B65G gene product [Bacillus subtilis]	50	27	750
910	2	438	887	gi 1104727	tyrosine-specific transport protein [Haemophilus influenzae]	50	25	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

UniProt ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
913	1	524	760	gi1205451	cell division inhibitor [Haemophilus influenzae]	50	32	237
973	1	424	236	gi1866947	orf3 gene product [Saccharomyces cerevisiae]	50	40	189
1009	1	653	429	gi1533727	M protein [group G streptococcus]	50	28	225
1027	1	511	257	gi1413934	ipa-10r gene product [Bacillus subtilis]	50	25	255
1153	2	556	326	gi1773676	nccA [Alcaligenes xyloxydans]	50	36	231
1222	1	798	400	gi1408485	1865G gene product [Bacillus subtilis]	50	21	399
1350	1	692	399	gi1289272	ferrichrome-binding protein [Bacillus subtilis]	50	32	294
2945	1	366	184	gi171704	hexaprenyl pyrophosphate synthetase (COI) [Saccharomyces erevisiae]	50	34	183
2968	2	1604	804	gi1397528	clumping factor [Staphylococcus aureus]	50	33	801
2998	2	657	394	gi1495698	F54E7.3 gene product [Caenorhabditis elegans]	50	40	264
3046	2	506	306	gi1513819 S138	acyl carrier protein - Anabaena variabilis (fragment)	50	32	201
3063	1	547	275	gi1474190	lucA gene product [Escherichia coli]	50	29	273
3174	1	3	146	gi1515900	alcohol dehydrogenase [Rhodobacter sphaeroides]	50	31	144
3792	1	625	314	gi1001423	hypothetical protein [Synecocystis sp.]	50	35	312
3800	1	2	262	gi144733	NAD-dependent beta-hydroxybutyryl coenzyme A dehydrogenase Clostridium acetobutylicum	50	28	261
3946	1	373	188	gi1576765	cytochrome b [Myrmecia pilosula]	50	38	186
3984	1	578	291	sp P37348 YECE	HYPOTHETICAL PROTEIN IN ASP5 5' REGION (FRAGMENT).	50	37	288
37	10	8250	7885	gi1204367	hypothetical protein (GB-U14003_278) [Haemophilus influenzae]	49	30	366
46	16	13802	14848	gi1466860	acd; B1308_P1_34 [Mycobacterium leprae]	49	24	1047
59	5	2267	3601	gi1606304	ORF_0462 [Escherichia coli]	49	27	1335
112	18	17884	18615	gi1559502	ND4 protein (AA 1 - 409) [Caenorhabditis elegans]	49	25	732
138	9	6973	7902	gi1303953	esterase [Acinetobacter calcoaceticus]	49	29	930
217	6	4401	5138	gi1496254	fibrinectin/fibrinogen-binding protein [Streptococcus pyogenes]	49	31	738
220	12	11803	12657	gi1397526	clumping factor [Staphylococcus aureus]	49	31	855
228	4	1842	2492	pir S23692 S236	hypothetical protein 9 - Plasmodium falciparum	49	24	651
268	1	5016	2614	gi1414047	ORF8 [Bacillus subtilis]	49	26	2403

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
271	2	1164	1373	gi 1001257	hypothetical protein [Synecocystis sp.]	49	38	210
300	3	4340	3180	gi 1510796	hypothetical protein (GPX91006_2) [Methanococcus jannaschii]	49	26	1161
381	1	2281	1142	gi 396301	metches PS00041: Bacterial regulatory proteins, arcC family signature [Escherichia coli]	49	29	1140
466	1	3	947	gi 1301863	Yqgp [Bacillus subtilis]	49	26	945
666	1	379	191	gi 633112	ORF1 [Streptococcus sobrinus]	49	29	189
670	2	403	1014	gi 1122758	unknown [Bacillus subtilis]	49	32	612
709	1	1433	795	gi 143830	xpaC [Bacillus subtilis]	49	29	639
831	1	943	473	gi 401786	phosphomannomutase [Mycoplasma pirum]	49	29	471
1052	1	422	213	gi 1303799	Yqen [Bacillus subtilis]	49	21	210
1800	1	342	172	gi 216300	peptidoglycan synthetase [Bacillus subtilis] sp P37585 HURG_BACSU HURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N-ACETYLURANYL-PENTAPEPTIDE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE RANSFERASE).	49	28	171
2430	1	2	376	ep P27434 YFCA_	HYPOTHEICAL 36.2 KD PROTEIN IN NDK-OCPE INTERGENIC REGION.	49	26	375
3096	1	542	273	gi 516160	surfactin synthetase [Bacillus subtilis]	49	25	270
32	4	3771	3100	gi 1217963	hepatocyte nuclear factor 4 gamma (HNF4gamma) [Homo sapiens]	48	36	672
38	1	1	609	gi 1205790	H. influenzae predicted coding region H1555 [Haemophilus influenzae]	48	28	609
45	6	5021	6427	gi 1524267	unknown [Mycobacterium tuberculosis]	48	20	1407
59	14	16346	31096	gi 11197336	LmpJ protein [Mycoplasma hominis]	48	28	14751
61	1	3	608	gi 1511555	quinolone resistance nraA protein [Methanococcus jannaschii]	48	30	606
61	3	3311	3646	gi 1303893	Yqhu [Bacillus subtilis]	48	29	336
114	1	98	415	gi 671708	su(s) homolog; similar to Drosophila melanogaster suppressor of abie (su(s)) protein, Swiss-Protein Accession Number P22293 [Drosophila virilis]	48	25	318
121	1	1131	610	gi 1314584	unknown [Sphingomonas S88]	48	29	522
136	1	2014	1280	gi 1205968	H. influenzae predicted coding region H1738 [Haemophilus influenzae]	48	23	735
171	10	8220	9557	gi 1208454	hypothetical protein [Synecocystis sp.]	48	34	1338
175	1	3625	1814	gi 396400	similar to eukaryotic Na+/H+ exchangers [Escherichia coli] sp P22703 VJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOMR-ACS INTERGENIC REGION (0549).	48	29	1812
194	1	2	385	gi 1510493	H. jannaschii predicted coding region M30419 [Methanococcus jannaschii]	48	25	384

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

CurIn ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
197	1	901	452	gi 1045714	ispermidine/putrescine transport ATP-binding protein [Mycoplasma genitalium]	48	25	450
203	1	1	396	gi 940288	protein localized in the nucleoli of pea nuclei; ORF; putative Pisum sativum	48	29	396
204	1	1363	698	gi 529202	No definition line found [Caenorhabditis elegans]	48	25	666
206	20	34815	27760	gi 511490	Gramicidin S synthetase 2 [Bacillus brevis]	48	27	7056
212	1	2	166	gi 295899	nucleolin [Xenopus laevis]	48	34	165
220	10	12652	11426	gi 44073	SecY protein [Lactococcus lactis]	48	23	1227
243	6	6450	5491	gi 1184118	malonate kinase [Methanobacterium thermoautotrophicum]	48	30	960
264	4	5434	3108	gi 1015903	ORF YJR151c [Saccharomyces cerevisiae]	48	26	2127
441	1	1332	768	gi 142863	replication initiation protein [Bacillus subtilis] pir B26580 B26580 replication initiation protein - Bacillus subtilis	48	23	765
444	5	3898	5298	gi 145836	putative [Escherichia coli]	48	24	1401
484	2	388	1110	gi 146551	transmembrane protein (kdp) [Escherichia coli]	48	18	723
542	3	1425	2000	pir 528969 5289	N-carbamoylascorine amidohydrolase [EC 3.5.1.59] - Arthrobacter sp.	48	27	576
566	1	3	1019	gi 153490	tetracycline C resistance and export protein [Streptomyces laevis]	48	24	1017
611	1	2	730	gi 1103507	unknown [Schizosaccharomyces pombe]	48	38	729
624	1	1255	665	gi 144859	ORF B [Clostridium perfringens]	48	26	591
846	1	1014	508	gi 537506	paranyosin [Dirofilaria immitis]	48	27	507
1020	1	66	950	gi 1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	48	30	885
1227	1	1	174	gi 493730	lipoygenase [Pisum sativum]	48	35	174
1266	1	1	405	gi 882452	ORF f211; alternate name ygaA; orf5 of X14436 [Escherichia coli] gi 41425 ORF5 (AA 1-197) [Escherichia coli] (SUB 15-211)	48	24	405
2071	1	707	381	gi 1408486	HS74A gene product [Bacillus subtilis]	48	25	327
2398	1	463	233	gi 1500401	reverse gyrase [Methanococcus jannaschii]	48	40	231
2425	1	476	246	pir H48563 H485	G1 protein - fowlpox virus (strain HP44) (fragment)	48	40	231
2432	1	446	225	gi 1353703	Trlo [Homo sapiens]	48	33	222
2453	1	794	399	gi 142850	division initiation protein [Bacillus subtilis]	48	29	396
2998	1	469	236	gi 577569	PepV [Lactobacillus delbrueckii]	48	31	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3042	1	14	280	gi 945219	mucin (Homo sapiens)	48	35	267
3686	1	1	405	gi 145836	putative [Escherichia coli]	48	25	405
4027	2	492	301	pir.S51177[S511	trans-activator protein - Equine infectious anemia virus	48	32	192
4	2	3641	2232	gi 1303989	YqkI [Bacillus subtilis]	47	24	1410
24	2	599	1084	gi 1540083	PC4-1 gene product (Bradyzia hyglida)	47	28	486
36	10	7524	6925	gi 1209223	esterase [Acinetobacter lwoffii]	47	26	600
43	2	196	1884	gi 1403455	unknown [Mycobacterium tuberculosis]	47	27	1689
44	22	116118	15108	gi 1511555	quinolone resistance norA protein protein [Methanococcus jannaschii]	47	31	1011
69	7	7141	6710	gi 438466	Possible operon with orfG. Hydrophilic, no homologue in the database; putative [Bacillus subtilis]	47	29	432
81	4	5022	4279	gi 466882	ppst: B1496_C2_189 [Mycobacterium leprae]	47	24	744
120	12	9135	8863	gi 927340	D9509.27p1 CAI: 0.12 [Saccharomyces cerevisiae]	47	38	273
142	1	2022	1174	gi 486143	ORF YKL094w [Saccharomyces cerevisiae]	47	32	849
168	3	2178	1093	gi 1177254	hypothetical BcsB protein [Bacillus subtilis]	47	29	1086
263	1	1884	943	gi 142822	D-alanine racemase cds [Bacillus subtilis]	47	34	942
279	1	1109	561	gi 516608	2 predicted membrane helices, homology with B. subtilis man Orf3 Rowland et. al. unpublished Accession number M7183), approximately 1 minutes on updated Rudd map; putative [Escherichia coli] sp P37355 VPB8_ECOLI	47	31	549
					HYPOTHETICAL 26.7 KD PROTEIN IN MEND-MENB			
345	2	2620	1676	gi 1204835	hippuricase [Haemophilus influenzae]	47	28	945
389	2	152	400	gi 456562	G-box binding factor [Dictyostelium discoideum]	47	32	249
391	1	1	831	gi 1420856	myo-inositol transporter [Schizosaccharomyces pombe]	47	19	831
404	3	2072	2773	gi 1255425	C33C8-2 gene product [Caenorhabditis elegans]	47	17	702
529	5	2145	3107	gi 1303973	YqjV [Bacillus subtilis]	47	29	963
565	2	2321	1257	gi 142824	processing protease [Bacillus subtilis]	47	28	1065
654	1	962	483	gi 243353	ORF 5' of ECRF3 [herpesvirus saimiri IVS, host-squirrel monkey, eptide, 407 aa]	47	23	480
692	1	115	633	gi 150756	40 kDa protein [Plasmid pJH1]	47	25	519
765	1	1634	819	gi 1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	47	28	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
825	2	211	1023	gi 397526	clumping factor (Staphylococcus aureus)	47	32	813
914	1	1	615	gi 558073	polymorphic antigen (Plasmodium falciparum)	47	29	615
1076	1	1	753	gi 1147557	Aspartate aminotransferase (Bacillus circulans)	47	33	753
1351	1	793	398	gi 755153	ATP-binding protein (Bacillus subtilis)	47	20	396
4192	1	3	293	gi 1145816	putative (Escherichia coli)	47	24	291
5	6	4708	4361	gi 305080	myosin heavy chain (Entamoeba histolytica)	46	30	348
11	4	2777	3058	gi 603639	Yel040p (Saccharomyces cerevisiae)	46	28	282
46	11	10518	10300	gi 1246901	ATP-dependent DNA ligase (Candida albicans)	46	28	219
61	4	3941	7930	gi 298032	EF (Streptococcus suis)	46	35	3950
132	4	5028	4093	gi 1511057	hypothetical protein SP:P45869 (Methanococcus jannaschii)	46	25	936
170	4	4719	3652	gi 1519105519	G4 protein - Sauroleishmania tarentolae	46	26	1068
191	7	9543	8284	gi 1041334	F5405.7 (Caenorhabditis elegans)	46	25	1260
253	1	1	396	gi 1204449	dihydrolyposamide acetyltransferase (Haemophilus influenzae)	46	35	396
264	3	437	973	gi 180189	cerebellar-degeneration-related antigen (CDR34) (Homo sapiens)	46	29	517
					cerebellar degeneration-associated protein (Homo sapiens)			
					pir A29770 A29770 cerebellar degeneration-related protein - human			
273	1	485	285	gi 607573	envelope glycoprotein C2V3 region (Human immunodeficiency virus type 1)	46	35	201
350	1	3	563	gi 537052	ORF_f286 (Escherichia coli)	46	35	561
384	1	2	862	gi 1221884	(urea7) amidolyase (Haemophilus influenzae)	46	31	861
410	4	1876	2490	gi 1110518	proton antiporter efflux pump (Mycobacterium smegmatis)	46	24	615
432	1	2663	1455	gi 1197634	orf4; putative transporter; Method: conceptual translation supplied by author (Mycobacterium smegmatis)	46	27	1209
458	1	2419	1211	gi 15470	portal protein (Bacteriophage SPPI)	46	30	1209
517	5	2477	4192	gi 1523812	orf5 (Bacteriophage A2)	46	23	1716
540	3	1512	1205	gi 215635	pacA (Bacteriophage P1)	46	30	228
587	2	649	1242	gi 537148	ORF_f181 (Escherichia coli)	46	29	594
1218	1	747	391	gi 1205456	single-stranded-DNA-specific exonuclease (Haemophilus influenzae)	46	30	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3685	1	1	402	gi1450688	hadM gene of EcoPRT gene product [Escherichia coli] pir[S38437 S38437 hadM protein - Escherichia coli pir[S05629 S05629 hypothetical protein A - Escherichia coli (S08 40-520)]	46	33	402
4176	1	673	338	gi1351460	PFM-C.1 gene product [Xenopus laevis]	46	31	336
37	7	4813	5922	gi1606064	ORF_408 [Escherichia coli]	45	24	1110
38	16	11699	12004	gi1452192	protein tyrosine phosphatase (PTP-BAS, type 2) [Homo sapiens]	45	24	306
87	2	1748	2407	gi11064813	homologous to sp:PHOR_BACSU [Bacillus subtilis]	45	23	660
103	12	14182	13385	gi11001307	hypothetical protein [Synecocystis sp.]	45	22	798
112	14	14791	13811	gi1204389	H. influenzae predicted coding region H10131 [Haemophilus influenzae]	45	23	981
145	4	4483	3461	gi1220578	open reading frame [Mus musculus]	45	20	1023
170	6	6329	4965	gi1238657	AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, epsilon, 514 aa]	45	27	1365
206	2	5230	4346	gi1222056	aminotransferase [Haemophilus influenzae]	45	27	885
228	1	60	716	gi1160299	glutamic acid-rich protein [Plasmodium falciparum] pir[A56514 A56514 glutamic acid-rich protein precursor - Plasmodium falciparum]	45	23	657
288	1	2	1015	gi11255425	G33C8.2 gene product [Caenorhabditis elegans]	45	23	1014
313	3	4339	3128	gi1581140	NADH dehydrogenase [Escherichia coli]	45	30	1212
332	1	914	459	gi1870966	P47A4.2 [Caenorhabditis elegans]	45	20	456
344	1	3	221	gi1171225	kinesin-related protein [Saccharomyces cerevisiae]	45	26	219
441	2	1501	1073	gi1142863	replication initiation protein [Bacillus subtilis] pir[B26580 B26580 replication initiation protein - Bacillus subtilis]	45	27	439
672	1	2	982	gi11511334	M. jannaschii predicted coding region MJ1323 [Methanococcus jannaschii]	45	22	981
763	3	1345	851	gi1606180	ORF_4310 [Escherichia coli]	45	24	495
886	3	379	846	gi1726426	similar to protein kinases and C. elegans proteins F37C12.8 and 37C12.5 [Caenorhabditis elegans]	45	30	468
948	1	3	473	gi1156400	myosin heavy chain (isoform unc-54) [Caenorhabditis elegans] pir[A93581 MMK myosin heavy chain B - Caenorhabditis elegans sp P02566 MYSB_CAEEL MYOSIN HEAVY CHAIN B (MHC B)]	45	25	471
1158	1	2	376	gi1441155	transmission-blocking target antigen [Plasmodium falciparum]	45	35	375
2551	1	4	285	gi11216705	ORF287 gene product [Porphyra purpurea]	45	28	282
3967	1	42	374	gi1976025	HraA [Escherichia coli]	45	28	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	7	6931	5846	gi1467378	unknown [Bacillus subtilis]	44	22	1086
138	8	6475	6849	gi173028	thioredoxin II [Saccharomyces cerevisiae]	44	28	375
221	5	7032	5617	gi153490	tetracycline C resistance and export protein [Streptomyces lauecensis]	44	21	1416
252	2	1331	1122	gi1204989	hypothetical protein [GB:000022.9] [Haemophilus influenzae]	44	30	210
263	2	3265	2093	gi1136221	carboxypeptidase [Sulfolobus solfataricus]	44	26	1173
365	4	4963	3524	gi1298822	orf1 gene product [Lactobacillus helveticus]	44	31	1440
543	3	1315	1833	gi1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]	44	24	519
544	4	3942	4892	gi1951460	PFM-C.1 gene product [Xenopus laevis]	44	32	951
792	1	1224	613	gi1205680	high molecular weight neurofilament [Rattus norvegicus]	44	28	612
44	18	11303	11911	gi1151614	molybdopterin-guanine dinucleotide biosynthesis protein A [Methanococcus jannaschii]	43	27	609
59	8	3665	5128	gi1153490	tetracycline C resistance and export protein [Streptomyces lauecensis]	43	21	1464
59	10	5516	7527	gi1153022	lipase [Staphylococcus epidermidis]	43	22	1992
99	1	1346	681	gi11419051	unknown [Mycobacterium tuberculosis]	43	21	666
310	8	9402	12134	gi1397526	clumping factor [Staphylococcus aureus]	43	21	2733
432	3	2782	2303	pir1A60540 A605	sporozoite surface protein 2 - Plasmodium yoelii (fragment)	43	29	480
519	3	2547	3122	sp1006530 DHSU_1.8.2.-1 (FC) (FCSD)	SULFIDE DEHYDROGENASE (FLAVOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR (EC 1.8.2.-1) (FC) (FCSD)	43	23	576
4	13	12053	13321	gi1295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]	42	18	1269
94	2	1768	1091	gi1501027	ORF2 [Trypanosoma brucei]	42	31	678
127	4	5791	4550	gi142029	ORF1 gene product [Escherichia coli]	42	21	1242
297	3	1515	1016	gi1142790	ORF1; putative [Bacillus firmus]	42	25	480
344	6	4097	3525	gi140320	ORF 2 (AA 1-203) [Bacillus thuringiensis]	42	30	573
512	1	2167	1115	gi1405957	yeeF [Escherichia coli]	42	23	1053
631	1	2434	1223	gi1580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir1506048 S06048 probable rodD protein - Bacillus subtilis ep131484 TAGE_BACSU PROBABLE POLY (GLYCEROL-PHOSPHATE) LPH-A-GUCCOSYLTRANSFERASE (EC 2.4.1.52) [TECHOTIC ACID BIOSYNTHESIS PROTEIN B]	42	24	1212

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
685	3	2359	1739	gi11303784	[YqeD [Bacillus subtilis]	42	19	621
4132	1	787	395	gi11022910	[protein tyrosine phosphatase [Dictyostelium discoideum]	42	25	393
86	2	1375	884	gi1309506	[spermidine/spermine N1-acetyltransferase [Mus saxicola] pir[S43430[S43430 spermidine/spermine N1-acetyltransferase - spiny ouse (Mus saxicola)	41	30	492
191	12	14797	14075	gi1124957	[orf4 gene product [Methanosarcina barkeri]	41	22	723
212	6	2150	3127	gi115873	[observed 35.2kd protein [Mycobacteriophage 15]	41	26	978
213	3	1263	2000	gi1633692	[TrsA [Yersinia enterocolitica]	41	18	738
408	4	2625	3386	gi11197634	[orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium smegmatis]	41	24	762
542	1	3	1103	gi1457146	[rhostry protein [Plasmodium yoelii]	41	21	1101
924	1	2	475	pir1JH0148[JH01	[nucleolin - rat	41	30	474
1562	1	1	402	gi1552184	[asparagine-rich antigen Pf35-2 [Plasmodium falciparum] pir[S27826[S27826 asparagine-rich antigen Pf35-2 - Plasmodium falciparum (fragment)]	40	20	402
2395	1	518	261	pir[S42251[S422	[hypothetical protein 5 - fowlpox virus	40	18	258
4077	1	3	305	gi11055055	[coded for by C. elegans cDNA YK171.5; coded for by C. elegans cDNA YK5C9.5; coded for by C. elegans cDNA YK1A9.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans]	39	21	303
958	1	1003	503	gi1255425	[C3308.2 gene product [Caenorhabditis elegans]	37	25	501
59	12	8294	10636	gi1535260	[STARP antigen [Plasmodium reichenowi]	36	24	2343
63	5	3550	8079	gi1298032	[BF [Streptococcus suis]	36	19	4530
544	3	2507	3601	gi11015903	[OHF YJR15ic [Saccharomyces cerevisiae]	35	22	1095
63	4	1949	3574	gi1552195	[circumsporozoite protein [Plasmodium falciparum] sp[P05691[CSP_PLAFL CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT)]	32	27	1626

TABLE 2

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4	1	1234	692
4	3	1712	2278
4	4	3703	3032
4	14	13073	12585
5	2	2539	1601
5	3	1532	1771
5	7	4741	4550
5	9	7939	6422
5	12	8711	8547
6	4	2359	1982
8	1	349	176
11	8	5144	5983
11	9	5968	6498
11	10	6472	6284
11	16	10954	11271
12	5	5352	4942
12	6	4596	4862
15	3	1895	1650
16	10	11263	10835
18	2	1093	917
20	9	9125	7764
20	10	8571	8230
20	112	9201	8803
20	13	12158	10470
23	1	674	339
23	6	6138	5485
23	8	6376	5942

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
23	9	7651	6881
23	15	12618	12830
24	4	4556	4185
24	6	5642	5241
25	2	1824	2402
31	2	505	849
31	3	1177	1524
31	4	2454	3005
32	2	765	1388
32	9	7952	8575
32	10	8591	8728
32	11	9718	9379
32	12	10797	10087
34	2	1315	1049
36	7	5226	5801
36	11	7575	7261
36	12	7424	7621
37	4	3158	2964
38	2	1585	980
38	11	6425	6868
38	20	16982	16171
38	26	20253	20804
38	27	20712	21264
39	1	1	627
40	1	805	404
43	1	796	428
44	4	2674	2124

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
44	5	2484	3263
44	14	10587	10129
44	20	13724	13536
44	21	13596	13994
45	7	6575	6297
46	8	6365	6520
46	12	10449	10976
46	17	15032	15424
47	1	288	1079
48	9	7620	7778
50	1	1612	962
50	2	1621	1316
51	1	738	370
51	5	2520	2245
53	1	442	287
53	7	6705	6319
54	7	9014	8709
55	1	592	326
55	3	1052	786
56	1	1	261
56	3	1551	1228
56	4	1970	1560
56	17	19092	18712
57	4	3694	3521
57	8	5436	5822
58	9	8885	8553
59	3	1366	1509

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
59	6	3026	2802
59	7	3770	3570
59	9	4946	4563
59	11	7518	8378
59	13	10401	16403
62	2	2696	1521
62	11	5440	5757
63	1	1	336
67	1	900	1781
67	2	1774	2610
67	3	2591	3904
67	8	7110	6955
68	1	78	326
70	6	6761	5199
70	11	8935	8645
77	3	1590	1192
79	2	1509	1228
79	3	1411	1791
83	1	2	403
85	9	8300	8653
85	10	8969	8781
86	3	1426	1232
87	8	9187	9366
88	3	1620	1922
89	1	3	161
89	7	5042	4878
91	1	1098	550

TABLE 3

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
91	3	3938	3141
92	2	449	928
92	3	1958	1467
92	9	5638	6024
94	1	661	332
94	3	2445	1813
94	4	2583	2197
96	11	10601	11050
99	6	4672	4523
99	7	5014	4784
100	8	7658	7287
102	7	4697	4368
103	3	2496	2035
104	1	2	694
104	2	699	1277
105	1	1235	693
105	3	3233	2655
105	1	3	221
106	3	1209	1355
107	1	1081	542
109	4	4025	3651
109	13	11625	11996
109	14	11981	12268
109	20	17401	17688
110	1	2	760
114	10	8764	9384
116	1	1	309

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
116	3	6273	4482
116	8	11049	9976
116	9	10313	10158
120	5	3703	3320
120	6	4270	3869
120	11	9290	9844
121	2	417	569
126	3	1090	818
127	3	2648	3196
127	5	4084	4395
131	6	6773	6438
132	2	715	1695
134	1	2	667
135	2	512	258
135	3	1124	729
138	1	3	152
138	7	6008	6463
140	1	2060	1032
140	2	2019	1513
140	5	2387	2743
142	2	1360	2388
142	7	8830	7586
143	7	7290	6502
144	1	1227	640
146	1	2	511
146	3	502	1350
146	4	3673	2540

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Config	ORF	Start	Stop
ID	ID	(nt)	(nt)
146	5	2874	3071
147	1	1	339
149	11	3956	3615
149	12	4036	3785
149	13	4507	4145
149	15	4807	4610
149	16	5495	5049
149	18	5739	5491
149	21	7416	7034
149	23	9216	8521
149	24	9681	9106
149	25	10679	9897
150	2	2303	1587
154	3	1795	1508
154	8	6586	6398
154	14	12704	12147
154	15	13531	12803
156	1	315	593
157	3	1183	2232
158	2	1471	1064
159	3	452	808
161	2	876	1808
161	6	4653	4279
161	7	4803	4540
161	8	4896	4717
161	11	5817	5638
163	2	1604	840

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
163	5	2796	2344
163	7	2952	2647
163	9	4905	5132
164	3	1338	1147
166	3	5213	4854
168	4	2500	2868
168	5	3595	4158
170	3	2517	2777
171	2	2277	1450
171	11	12576	11125
172	1	3	278
172	2	1940	1149
173	1	1289	708
173	5	7001	6114
174	2	593	1105
175	3	2552	2890
175	5	3820	3335
175	7	4362	4506
182	4	5477	4986
184	5	6043	5702
188	2	1210	1755
188	4	2647	2994
189	6	2614	3039
190	3	1998	2564
191	1	1	153
191	2	950	669
191	10	11786	13039

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
191	11	12902	12363
192	1	91	426
195	3	2306	1932
195	5	2899	2606
198	2	1016	1591
201	1	170	625
203	2	783	1466
206	6	8930	7815
206	12	13947	13636
206	21	28208	27960
212	2	170	817
212	3	796	1167
212	7	3128	3436
212	9	3749	4075
213	1	1	705
214	2	1076	570
214	6	4084	3738
214	9	6600	6995
214	10	7864	7469
217	1	1927	965
218	1	178	657
218	3	1776	2156
220	2	1851	1369
220	3	3251	2262
220	7	8275	7208
220	8	10244	8661
220	9	11796	10216

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
221	4	3095	2613
221	9	11428	110757
226	1	3	659
226	2	2196	1459
226	3	1476	1961
227	1	2	487
227	2	460	975
227	4	1855	2121
227	5	2052	2345
227	6	4760	3768
227	9	5591	6367
228	5	2503	2877
228	6	2846	3526
233	7	3944	3762
236	2	809	579
238	2	1975	1391
239	2	1417	905
241	5	4495	4314
242	2	1677	1363
243	1	127	576
244	1	1291	647
244	2	3035	1962
245	2	1614	1258
246	1	69	215
246	4	738	1733
249	3	3906	3712
250	1	494	249

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
254	1	1	156
256	2	956	1144
257	3	3700	3227
260	4	4906	4580
261	4	2196	2606
261	6	3214	3681
264	2	155	439
264	5	5252	4533
264	6	4739	5107
267	2	1323	931
268	4	5140	4700
272	1	862	446
272	3	1200	1439
272	9	4691	4909
272	10	6469	6035
276	4	1746	1901
278	1	224	553
278	5	3299	3448
278	7	4849	5127
285	2	551	736
288	3	1756	1950
288	5	2055	2276
289	1	2107	1055
290	2	2234	1932
291	2	332	622
291	5	1545	2051
295	3	1606	1349

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
295	4	2728	2141
295	5	2220	2762
297	2	788	465
298	1	2	205
300	2	2180	1928
301	7	2794	2624
304	1	3	194
306	1	109	654
306	5	4036	4257
307	1	674	339
307	8	3645	3995
308	1	1	654
308	2	1120	599
308	4	2643	2332
313	2	2314	1919
314	1	10	702
316	2	982	1341
316	6	2758	3165
317	1	2	1114
317	3	4570	3458
321	6	5645	5217
321	7	6319	6140
321	8	7450	6794
322	2	827	543
326	2	165	1112
326	3	1117	1467
328	1	936	469

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
328	5	3452	3276
329	1	3	719
329	2	781	1212
329	3	1471	1833
330	1	576	289
330	3	1447	1623
332	3	2353	2204
332	7	4971	5138
333	2	3295	3128
335	1	864	433
337	2	95	526
340	2	1658	1356
341	1	3	281
341	3	2476	3192
341	5	3618	3944
341	6	3929	4558
344	5	3197	2889
345	1	1532	768
346	2	221	592
350	3	1410	1598
352	2	2178	1765
352	3	7316	4596
352	7	7967	8404
352	8	8906	9247
352	9	10171	9854
359	1	1	346
362	1	3	656

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
364	2	2158	1808
364	8	10974	10714
365	2	1612	1313
365	5	4680	4090
365	7	4980	6219
366	3	520	1719
367	3	906	1085
368	1	748	494
375	1	2	136
380	3	1351	1097
389	1	1	276
390	1	2	877
390	2	1173	1549
391	2	751	560
395	1	391	197
396	1	2132	1068
398	3	1344	1141
399	1	176	669
401	3	566	847
402	2	100	465
404	8	5561	5370
408	2	3507	2269
408	3	2875	2672
408	5	3524	4423
410	3	2111	1890
413	1	880	488
416	1	607	320

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
416	2	578	847
416	3	2195	1590
417	1	3	179
417	2	161	616
420	2	788	513
422	2	357	677
431	2	856	1407
432	2	446	1084
433	1	1	417
433	3	2311	2033
434	1	942	535
434	2	2089	1235
440	1	1	450
442	2	1269	3320
443	3	1873	1520
444	1	1	696
444	7	6761	6366
451	1	940	614
453	2	896	636
453	8	3833	4786
453	9	4718	4512
453	10	4937	4731
455	1	434	219
455	2	472	930
459	1	265	687
462	1	2	247
466	2	1494	907

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
467	1	654	349
468	1	2	250
469	1	1488	925
469	3	2386	3372
469	4	3464	3706
470	1	77	538
470	6	4098	3694
470	7	6330	5886
470	9	7351	8181
470	10	8175	9773
471	1	940	500
471	2	1562	1017
476	1	70	267
477	1	2	760
477	3	1764	2081
477	4	2066	2332
480	5	4016	4261
481	2	956	480
486	3	613	774
487	6	1795	2112
488	1	715	359
492	1	127	675
493	1	2	520
493	2	496	1242
502	3	1149	1571
504	1	690	346
505	5	4566	4150

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
511	2	1741	1232
512	2	583	747
515	1	609	812
517	4	2179	2511
520	4	2097	2360
520	6	3908	3669
527	1	1	498
528	1	637	335
529	2	1679	1104
530	7	5298	5534
536	1	308	156
538	1	1362	736
538	3	2203	2880
538	5	3531	3121
538	6	4348	3731
540	1	996	664
540	2	1495	1031
541	1	89	433
541	2	719	432
542	2	1048	1272
545	2	1012	734
551	1	2145	1129
555	2	892	704
558	3	1357	1154
558	4	1760	1458
558	5	2105	1821
558	6	2166	2020

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
558	7	2636	2322
558	8	3053	2802
558	9	3986	3453
560	1	475	921
565	3	1706	1485
571	1	308	156
571	3	994	1206
577	1	2	199
577	2	163	453
579	1	1	477
579	2	1784	1200
583	1	1988	996
585	1	946	539
587	1	22	573
588	2	1896	1372
588	3	1742	1554
590	1	47	334
592	2	1455	1141
593	1	2	775
593	2	817	1122
595	1	87	890
596	3	1593	1435
602	1	8	169
603	5	1071	1469
606	1	322	768
607	5	1444	1226
610	1	1029	541

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
612	1	3	500
616	1	991	650
617	2	716	491
622	1	36	347
625	4	2046	2549
627	1	67	210
628	1	901	452
631	3	4789	4004
634	1	1448	759
636	1	189	368
636	2	1929	1063
637	2	2323	1994
638	1	227	1081
639	1	518	261
639	2	1377	811
641	1	118	444
642	3	1615	1331
642	4	2260	1847
643	1	3	608
645	4	1534	1758
645	6	2025	2121
645	7	2940	2488
648	1	2	1045
660	1	77	601
660	2	576	872
661	1	1725	961
664	2	89	304

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
667	1	3	413
668	1	1	330
671	2	812	916
673	1	3	338
674	2	865	984
679	1	1	237
679	3	1589	1906
688	1	1236	835
688	2	1352	1077
694	1	3	143
696	2	818	432
706	1	36	224
709	3	1183	1449
711	1	3	908
715	1	3	167
716	1	2	637
721	1	133	570
722	1	763	383
723	1	1656	829
723	2	1498	1112
727	1	2	472
729	1	268	441
731	1	130	828
735	1	2	214
736	1	3	782
738	1	2	298
742	1	3	230

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
745	3	1148	760
748	2	282	464
749	1	685	344
751	1	901	452
755	1	97	522
755	2	520	918
758	2	663	400
764	2	1033	746
767	1	1	405
768	1	2	373
771	1	1058	534
778	1	1735	902
785	1	1790	1023
787	1	1260	631
791	1	3	224
799	1	15	260
804	1	304	711
805	1	3	680
808	1	219	842
810	1	2221	1112
810	2	1774	1442
812	1	38	979
817	1	714	358
818	2	487	1104
819	2	1529	1032
819	3	1748	1419
820	1	195	1064

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
828	1	506	255
829	1	48	800
830	1	578	291
832	1	594	298
835	1	370	796
840	3	491	709
845	1	912	457
850	2	303	449
853	1	715	359
860	1	2	256
864	1	18	410
868	2	383	715
864	6	1676	1828
870	1	1	588
873	1	906	454
875	1	584	294
877	1	1661	1020
878	1	981	544
879	1	1567	785
881	1	1	243
882	1	389	604
890	1	2	508
905	1	793	398
906	1	852	544
912	1	373	188
913	1	3	290
913	2	1092	547

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
915	1	6	161
915	2	169	402
921	1	126	386
927	1	1578	808
928	1	2	385
929	1	2	400
932	1	2	400
934	1	1	384
936	1	1052	528
937	1	2	616
945	1	220	645
945	2	649	1242
946	1	1702	950
949	1	1	270
951	1	3	362
955	1	3	143
960	1	723	400
963	1	1	162
965	1	690	346
966	1	1079	606
969	1	3	302
971	1	12	170
974	1	319	161
976	1	692	348
977	1	2	211
982	1	1926	982
984	1	589	296

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
987	1	3	467
993	1	1	525
994	1	920	549
1004	1	557	318
1014	1	624	313
1015	1	2	463
1016	1	288	145
1019	1	1205	660
1022	1	839	474
1024	1	595	299
1024	2	276	431
1030	1	673	338
1032	1	355	179
1040	1	794	399
1043	1	3	269
1044	2	115	399
1047	1	1	159
1051	1	704	354
1051	2	1233	733
1063	1	2	400
1069	1	2	148
1069	2	769	533
1075	1	707	399
1077	1	97	405
1081	1	58	438
1086	1	1	384
1087	2	246	431

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1088	1	3	174
1096	1	474	238
1098	1	1015	509
1100	1	1020	511
1100	2	1520	1158
1101	1	703	353
1102	1	385	194
1107	1	2	580
1114	1	3	422
1115	1	2	268
1119	1	22	267
1129	1	40	342
1132	1	360	141
1133	1	609	376
1148	1	446	225
1147	1	558	280
1153	1	1	153
1154	1	3	818
1159	1	1	330
1161	1	341	186
1164	1	427	254
1171	1	19	240
1171	2	108	299
1183	1	2	379
1195	1	355	179
1196	1	1	189
1200	1	33	197

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1203	2	129	464
1222	2	105	401
1232	1	1	387
1240	1	2	175
1247	1	520	311
1271	1	412	221
1286	1	2	595
1295	1	1	165
1306	1	367	185
1314	2	158	631
1316	1	58	570
1359	1	384	192
1370	1	1	402
1371	1	1	345
1374	1	710	357
1378	1	2	400
1392	1	3	413
1411	1	202	432
1433	1	331	167
1450	1	2	256
1453	1	295	149
1471	1	721	398
1477	1	869	639
1502	1	794	399
1518	1	126	449
1534	1	283	143
1546	1	3	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1547	1	506	255
1583	1	3	350
1587	1	3	563
1602	2	170	679
1629	1	1	402
1665	1	468	235
1760	1	625	314
1762	1	3	200
1876	2	119	286
1895	1	2	379
1931	1	798	400
1976	2	715	383
2055	2	253	401
2056	1	331	167
2150	1	523	263
2157	1	794	399
2164	1	564	283
2175	1	218	400
2212	1	492	331
2338	1	732	367
2342	1	3	167
2352	1	330	166
2352	2	622	398
2355	1	47	352
2356	1	679	341
2359	1	301	352
2421	1	296	150

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1046	1	167	185
1049	1	553	278
1050	1	3	314
1052	1	504	253
1065	1	2	157
1070	1	357	190
1075	1	440	222
1080	1	1	285
1092	1	320	162
1093	1	411	250
1100	1	52	237
1103	1	47	298
1118	1	344	174
1123	1	2	145
1127	1	1	147
1138	1	336	169
1142	1	388	203
1144	1	664	386
1151	1	337	170
1155	2	202	384
1168	1	12	176
1205	1	288	145
1282	1	1	150
1303	2	239	400
1371	2	211	399
1358	1	2	148
1358	2	36	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3118	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

5. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3568	1	751	377
3595	1	757	380
3618	1	2	338
3618	2	110	402
3622	1	86	358
3622	2	664	398
3642	1	876	439
3649	1	781	398
3651	1	625	314
3664	1	467	637
3674	1	55	402
3677	1	619	311
3704	1	1	402
3726	1	535	269
3765	1	510	256
3779	1	554	357
3794	1	266	135
3794	2	667	377
3796	2	638	375
3801	1	474	262
3806	1	453	298
3807	1	42	389
3815	1	798	400
3827	1	3	320
3842	1	781	392
3853	1	671	399
3855	1	1	324

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3857	1	2	235
3861	1	590	297
3865	1	695	399
3897	1	3	173
3897	2	143	400
3898	2	225	401
3921	2	103	342
3927	1	70	375
3930	1	76	234
3946	2	651	382
3951	2	105	377
3965	1	646	344
3973	1	795	400
3981	1	3	311
3998	1	3	356
4001	1	481	296
4003	1	90	335
4018	1	2	259
4018	2	186	401
4021	1	1	345
4043	1	3	344
4054	1	3	344
4066	1	1	150
4070	1	1	324
4072	2	187	390
4073	1	1	285
4077	2	127	372

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4083	1	3	359
4090	1	27	368
4101	1	103	297
4105	1	1	306
4107	1	570	286
4119	1	629	339
4121	1	740	372
4123	1	3	230
4127	1	3	341
4128	1	2	331
4130	1	768	415
4166	1	97	381
4157	1	3	206
4186	1	505	254
4224	1	510	256
4239	1	1	348
4242	1	709	356
4252	1	589	296
4253	1	1	174
4256	1	568	323
4258	2	498	334
4267	1	284	144
4271	1	2	304
4287	1	303	163
4289	1	471	319
4302	1	153	305
4304	1	1	186

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4304	2	96	314
4306	1	2	151
4318	1	576	289
4322	1	5	148
4331	1	439	221
4331	2	528	364
4338	1	728	399
4346	1	471	277
4367	2	117	311
4373	1	2	268
4381	1	574	326
4384	1	614	309
4397	1	9	311
4402	1	1	249
4403	1	606	328
4406	1	3	317
4411	1	2	280
4411	2	697	398
4412	1	2	364
4418	1	3	230
4424	1	601	398
4443	1	427	215
4471	1	643	323
4478	1	540	271
4482	1	50	289
4489	1	601	302
4491	2	12	206

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4495	1	3	179
4496	1	500	252
4500	1	130	306
4511	1	493	248
4518	1	1	246
4526	1	480	241
4527	1	2	163
4532	1	3	239
4542	1	11	175
4567	1	36	200
4573	1	1	231
4578	1	642	322
4619	1	1	180
4620	1	349	176
4662	1	1	246
4669	1	2	157
4680	1	28	183
4690	1	344	174

Table 4

ORF	SEQ ID NO	BLAST HOMOLOG	Antigenic Regions			
			Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein	36-45	84-103	152-161	176-185
238_1	5193	chrA	21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub)	20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1	20-29	59-73	85-97	162-171
276_2	5196	lipoprotein	21-33	65-74	177-186	211-220
45_4	5197	ProX	28-37	59-69	85-100	120-129
316_8	5198	hypothetical protein	45-54	88-97	182-192	243-253
154_15	5199	unknown	31-40	48-58	79-88	95-104
228_3	5200	unknown	25-38	40-52	64-74	80-89
228_6	5201	unknown	29-41	89-101	128-143	173-184
50_1	5202	unknown	21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic	21-31	58-67	92-101	111-120
442_1	5204	unknown	30-39	91-100	122-137	182-192
66_2	5205	unknown	50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic	19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein	27-36	86-95	129-138	192-201
161_4	5208	SphX	27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)	21-33	61-70	83-92	100-109
942_1	5210	traH [Plasmid pSK41]	83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)	12-22	87-96	111-120	151-160
20_4	5212	peptidoglycan hydrolase (S)	24-34	129-138	141-150	161-171
328_2	5213	lipoprotein (H. flu)	81-90	123-133	290-299	
520_2	5214	fibronectin binding protein	44-54	63-79	81-90	95-110
771_1	5215	emm1 gene product (S. pyo)	30-39	65-82	96-106	112-121
999_1	5216	predicted trithorax prot. (D	7-16	120-129	157-166	
853_1	5217	ORF2136 (Marchantia polyr	43-52	88-97	102-111	
287_1	5218	psaA homolog	13-22	28-44	72-82	114-124
288_2	5219	cell wall enzyme	14-23	89-98		
596_2	5220	penicillin binding protein 2b	40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindi	28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp	10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prote	4-13	29-47	60-73	90-99
171_11	5224	EF	20-31	91-110		
63_4	5225	penicillin binding protein 2b	12-21	59-68	95-104	
353_2	5226		46-55	62-71		
743_1	5227	29 kDa protein in fimA regi	23-32	68-79	94-103	175-184
342_4	5228	Twitching motility	10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein	97-106	132-141	158-167	180-189
70_6	5230	nodulin	36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie	8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)	26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure	72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic	24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)	59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. au	45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisae	76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure	22-32	71-80	89-98	114-122
87_7	5239	5' nucleotidase precursor ('	29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub	102-111			

Table 4

5	ORF	Antigenic		Regions			(cont)
		Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
	168_6	244-272	303-315				
	238_1	260-269	291-301	308-317			
	51_2	140-152	188-208	211-220	256-266	273-283	
10	278_3	198-209					
	276_2	255-268					
	45_4	177-199	221-230	234-243	268-279	284-293	304-313
	316_8						
	154_15	148-157	177-187	202-211			
15	228_3	101-119	139-154	166-181			
	228_6						
	50_1						
	112_7	136-149	197-211	218-229	253-273		
	442_1	199-210	247-257	264-277	287-309		
20	66_2						
	304_2	178-187	250-259				
	44_1						
	161_4						
	46_5	131-141	162-176	206-215	243-252	264-273	285-294
25	942_1						
	5_4	189-205	230-239	246-264	301-318	340-354	378-387
	20_4	202-212	217-234	260-275	314-336	366-373	380-391
	328_2						
	520_2						
30	771_1	145-154					
	999_1						
	853_1						
	287_1	154-164					
	288_2						
35	596_2	121-130					
	217_5	244-253	259-268	288-297	302-311		
	217_6	144-158	174-183	188-197	207-216	226-242	
	528_3						
40	171_11						
	63_4						
	353_2						
	743_1	197-207					
	342_4						
45	69_3	195-211					
	70_6	206-215	263-272	291-301	331-340	358-371	390-414
	129_2	117-127	141-157	168-183	202-211	222-231	261-270
	58_5	184-203	260-269	275-299	330-344	372-381	424-433
	188_3						
50	236_6	138-147	163-172	187-198	244-261	268-278	308-317
	310_8	131-140	144-153	177-186	190-199	204-213	216-227
	601_1	208-218					
	544_3	170-179	184-193	224-235	274-287	327-336	352-361
	662_1						
55	87_7						
	120_1						

Table 4

5	ORF	Antigenic		Regions	(cont)		
		Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5	306-315					
25	942_1						
	5_4	393-407	416-426	456-465			
	20_4	396-405	410-419	461-481			
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6	453-471	506-515				
	129_2	296-315					
	58_5						
	188_3						
50	236_6	358-377	410-423	428-439	442-457	467-476	480-493
	310_8	238-251	256-275	281-290	296-310	314-333	338-347
	601_1						
	544_3						
	662_1						
55	87_7						
	120_1						

Table 4

5	ORF	Antigenic Regions (cont)				
		Region 17	Region 18	Region 19	Region 20	Region 21
	168_6					
	238_1					
	51_2					
10	278_3					
	276_2					
	45_4					
	316_8					
	154_15					
15	228_3					
	228_6					
	50_1					
	112_7					
	442_1					
20	66_2					
	304_2					
	44_1					
	161_4					
	46_5					
25	942_1					
	5_4					
	20_4					
	328_2					
	520_2					
30	771_1					
	999_1					
	853_1					
	287_1					
	288_2					
35	596_2					
	217_5					
	217_6					
	528_3					
	171_11					
40	63_4					
	353_2					
	743_1					
	342_4					
	69_3					
45	70_6					
	129_2					
	58_5					
	188_3					
	236_6					
50	310_8	357-366	370-379	429-438	443-452	478-487
	601_1					551-560
	544_3					
	662_1					
55	87_7					
	120_1					

Table 4

	ORF	Antigenic Regions		(cont)			
		Region 23	Region 24	Region 25	Region 26	Region 27	Region 28
5	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5						
25	942_1						
	5_1						
	20_4						
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6						
	129_2						
	58_5						
	188_3						
	236_6						
50	310_8	622-632	670-685	708-718	823-836	858-867	877-886
	601_1						
	544_3						
	662_1						
	87_7						
55	120_1						

Table 4

5

10

15

20

25

30

35

40

45

50

55

ORF	Antigenic Regions		(cont)
	Region 29	Region 30	
168_6			
238_1			
51_2			
278_3			
276_2			
45_4			
316_8			
154_15			
228_3			
228_6			
50_1			
112_7			
442_1			
66_2			
304_2			
44_1			
161_4			
46_5			
942_1			
5_4			
20_4			
328_2			
520_2			
771_1			
999_1			
853_1			
287_1			
288_2			
596_2			
217_5			
217_6			
528_3			
171_11			
63_4			
353_2			
743_1			
342_4			
69_3			
70_6			
129_2			
58_5			
188_3			
236_6			
310_8			
601_1			
544_3			
662_1			
87_7			
120_1			

Table 4

ORF	BLAST		Antigenic Regions			
	HOMOLOG		Region 1	Region 2	Region 3	Region 4
46_1	5241	aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242	glycerol ester hydrolase (P.	9-26	57-73	93-107	123-133
174_6	5243	ketopantoate hydroxymeth	71-80	203-212	242-254	265-274
206_16	5244	ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245	NaH-antiporter protein (E. f	120-129	332-347	398-408	
322_1	5246	acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247	transport ATP-binding prote	108-126	218-227	298-308	315-334
214_3	5248	2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249	clumping factor	5-14	43-54	59-68	76-95
685_1	5250	signal peptidase	59-68	72-81	86-95	99-108
54_3	5251	fibronectin binding protein	23-32	37-46	50-59	89-98
54_4	5252	fibronectin binding protein	43-52	66-75	95-104	147-156
54_5	5253	fibronectin binding protein	49-60	81-90		
54_6	5254	fibronectin binding protein	55-71	82-97	139-158	175-186
328_1	5255	lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
46_1	215-242	333-352	376-385	416-432	471-487	
63_4	145-154	191-202	212-223	245-265	274-283	291-300
174_6						
206_16	239-259	275-284				
267_1						
322_1	298-319	350-359				
415_2	344-353	371-380	395-404	456-465	486-495	518-527
214_3	318-337	365-375				
587_3	106-115	142-151	156-166	173-182	186-198	204-213
685_1	113-122	130-145				
54_3	128-138	185-194	217-226	251-260	268-277	295-305
54_4	175-188	191-200	203-212	220-229		
54_5						
54_6	220-230	287-304	317-326	344-353	364-373	378-387
328_1						

Table 4

ORF	Antigenic Regions		(cont)			
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
46_1						
63_4	306-315	319-328	366-376	395-420	453-462	467-476
174_6						
206_16						
267_1						
322_1						
415_2	539-555					
214_3						
587_3	217-226	278-287	318-327	332-342	351-360	377-386
685_1						
54_3	316-325	329-345	355-372	387-396	416-425	438-448
54_4						
54_5						
54_6	396-407	427-436	514-531	541-550	569-578	612-622
328_1						

Table 4

ORF	Antigenic Regions		(cont)			
	Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
46_1						
63_4	485-500	513-525				
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	396-405	426-442	459-470	485-494	505-514	531-562
685_1						
54_3	455-462	472-491	517-536			
54_4						
54_5						
54_6	639-648	673-681	703-715	723-732	749-760	772-788
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 24	Region 25	Region 26	Region 27	Region 28	Region 29
46_1						
63_4						
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	567-578	584-601	607-840	844-854	858-870	877-886
685_1						
54_3						
54_4						
54_5						
54_6	793-802	811-826	834-848	866-876	893-903	907-918
328_1						

Table 4

ORF	Antigenic Regions (cont)	
	Region 30	Region 31
46_1		
63_4		
174_6		
206_16		
267_1		
322_1		
415_2		
214_3		
587_3	889-911	927-936
685_1		
54_3		
54_4		
54_5		
54_6	925-944	951-997
328_1		

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: US
- (F) POSTAL CODE: 20850

(ii) TITLE OF INVENTION: Staphylococcus aureus Poly-nucleotides and Sequences

(iii) NUMBER OF SEQUENCES: 5255

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
- (B) COMPUTER: HP Vectra 486/33
- (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/009,861
- (B) FILING DATE: 05-JAN-1996

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5895 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

10 TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTTT TAAAATCCCA 60
 GTAATAAAAT CAAAAAATAA GTTAAATAAT GTATTCATTT TAAGTCCTCC TTAATAAAGa 120
 15 aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG 180
 GCTCTTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT AACATAATC GCATAATATT 240
 TTTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA 300
 20 CTGAGATTAC ACCTAAAGAA ATAAGTGTAA AAATAATCAT AATTAAAAAG TTAATATGAA 360
 AATTTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA 420
 AAGAAGAAGG TGCATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTCG 480
 25 TAATTAACGT AAACAGAAAA ATGTTTAAAA TATAGGCAAT AGTATACATA ACAATTAATT 540
 TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAAT TGAATATGAT 600
 CTAAGTTATT TCTCTTTTGA AGATACGTGG CAAACTGGTC AATTTTATTA TCAAAATAAT 660
 TCAATTTTAC ACCACTCTCC TCACTGTGCAT TATACGATTT AGTACAATCT TTTATCATTa 720
 TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTAAAT GTAATTGTGA TTTAATATTT 780
 35 TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAT 840
 AAATGCTTTT AGCATGTTTT AATATACTA GATCACAGAG ATGTGATGGA AAATAGTTGA 900
 TGAGFTGTTT AATTTTAAGA ATTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC 960
 40 AAGATATCAT TTCAACAATC GGTGACTTAG TAAAATGGAT TATCGACACA GTGAACAAAT 1020
 TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTGTGTA TCTTCGTATA 1080
 GTAATAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT 1140
 45 GAGTAACATT TGCTAGTAGA GTTAGTTTCC TTGGACTCAG TGCTATGTAT TTTTCTTAAT 1200
 TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT 1260
 50 TATTTGATAA ATAAAATTTT TTTCATAATT AATAACATCC CCAAAAATAG ATTGAAAAAA 1320
 TAACTGTAAA ACATTCCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC 1380
 CTTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAATAT 1440

	TCATTTGCAA AGGGCGAAAT GGGTTCTTAC TGAGTTATCT ATTATAAAAA AATAAACATA	1560
	GACTTATGAA AAATCTCTCA TAAATCTATG TTTAGTCATG aCATGTGTTA AATATTATTT	1620
5	CGGGCGCTTC TTATTTATAC AAATCTAATT TAATACTTTT AAATACAGGT ATATTTTCgC	1680
	GTTGCTGTTC TACTTCATTT AAGTTTAAAT CTACAGTCAA AATATCTGCG GATTTCATTTA	1740
10	ATTCTCCAAC TAAATCTCCA TTTGGGTTTA TAACTATCGA ATGACCAGCA TATTCTGTGT	1800
	TACCATCGAA TCCAGTGCTA TTAGTTCCAA TGACAAACAT ATTATTTTCA ATTGCACGTG	1860
	CCTTTAGTAA TGAATGCCAA TGTGAAGAC GTGACATAGG CCATTGCGCC ACATAAAATG	1920
15	CAATTTTAGC ACCACTACGA GCAGGATATC TTAATAATTC TGGAAAACGT AAATCATAAC	1980
	AGATAAGTTG GGTCACATAA GTACCGTCAG ACAATTGAAA GGGTTCAGCT ACGTATTCGC	2040
	CAGCGTTAA AAATTCATGC TCTCTTAACA TAGGAACTAA ATGAACTTTG TCGTATTCaT	2100
20	TAATCAGCTG GCCACTTTTA TTCACACTAA AAGCTGTATT AAATATTTGA TTGTTTCTAA	2160
	TGTTAGAAAC TGACCCAGCT ACGATATCGA CTTTATATTT TTCAGCTAAA TGTTTAATAA	2220
	ATGAAAACT TTGTCCTAGA TTATTATCTG CTTTTTCATT TAAATGCTCT AAATCATAGC	2280
25	CATTATTCCA CATTTCAGGT AAAACGACTA CATCTACTTC AGCATTGATA TTTTTTTCGA	2340
	ACCATTGCGT TATTTGAGTT TCATTTTTAG AACTATCTCC AAAAACAATC GGTAATTGAT	2400
30	AAATTTGGAC TTTCATAACA TCACATCCTT GATAGATCTT ATATATAACT TACTAAAAGT	2460
	TATGTTGAAA CGCAAAAAAC GAGCACAAGA CATAAAATCA AAGTCCTAGG CTCTACAAAG	2520
	TTATATTGAC AGTAGTTGAT GGGGCCCCAA CATAGAGAAA TTGGAACACC AATTTCTACA	2580
35	GACAATGCAA GTTGGGGTGG GCTCTAACAT AAAGAAATAC TTTTCTTTA GAAATTAGTA	2640
	TTTCTTATAC ATGAGTTTTA CTCATGTATT CCTATTCTTA AGTGCACATT AGCAGCGGCT	2700
	AATGTGTAAG AACTACTACA TAATGAATAA CTAATGATTC TTTATCATTT CTGTCCCATT	2760
40	CCTAACAATA TATTGATTAT TTTTTTATTA CGAAACGATC TTCCACTGGA TTAAATGTTT	2820
	TTTCGCCAGC AGCTTCACGA ATATCACCAA ATGGCATTG AGCAATAAGT TTCCAACTTT	2880
45	TAGGAATATT AAATTCATTT GAAGTCATCT CATCAACAAG TGGATTATAG TGTTGTAATG	2940
	AAGCACCTAT GCCTTTAGTA GCTAATGCAG TCCAAATTGC AAATTGATGC ATGGCATTG	3000
	TTTGAGTTGA CCATATTGCA AAATTATCAT AGTAGTTTGG CATTGTGTTCT TGTAACCAC	3060
50	TTACAACATC TTGATCTTCA TAAAACAAAA TTGTACCGTA TGAATGTTG AAGTTATCAA	3120
	TTTTTTGTTC AGTTGGCTCG AAATCACGAT TCTCTCCCAT GACTTCTTTT AAAATTGCTT	3180
55	TTGTGTTATC CCAAATTTA TTATTGTTGT CATTTAACAA GAGAACAATT CTAGTTGATT	3240

	CATCGCTAAT	TGATATCGAA	TCTTTCAAAT	TATATATTGA	ACGTCTTTCT	TCCATTGCAT	3360
	TGTCAAAAGT	CATTGCTTTT	TTATCTTTTT	TAAATAAGCC	CATAATTATT	GTCCTTCTT	3420
5	TAGTAAAGAA	TACTTAATAG	ACTAAGTATA	AAATTTATAC	TCGTACTTGT	AAAGCAATAT	3480
	TTACGAAAAT	TTCAAGAATA	TTAATATTCA	TTTTCAAATT	CCAAATATAA	ATGCATTTTC	3540
10	AACGCATATT	TATTATACTT	AGATTAATAC	TTACATGAAA	AAGGGAGGTG	TCTCGTGAAA	3600
	TGTCATATCA	TTGGTTTAAG	AAAATGTTAC	TTTCAACAAG	TATTTTAATT	TTAAGTAGTA	3660
	GTAGTTTAGG	GCTTGCAACG	CACACAGTTG	AAGCAAAGGA	TAAGTTAAAT	GGAGAAAAAC	3720
15	CAACTACTAA	TTTGAATCAT	AATATAACTT	CACCATCAGT	AAATAGTGAA	ATGAATAATA	3780
	ATGAGACTGG	GACACCTCAC	GAATCAAATC	AAACGGGTAA	TGAAGGAACA	GGTTCGAATA	3840
	GTCGTGATGC	TAATCCTGAT	TCGAATAATG	TGAAGCCAGA	CTCAAACAAC	CAAAACCCAA	3900
20	GTACAGATTG	AAAACCAGAC	CCAAATAACC	AAACTCAAG	TCCGAATCCT	AAACCAGATC	3960
	CAGATAACCC	GAAACCAAAA	CCGGATCCAA	AACCAGACCC	AGATAAACCA	AAGCCAAATC	4020
25	CGGATCCAAA	ACCAGATCCA	GATAACCCGA	AACCAAATCC	AGATCCAAAA	CCAGACCCAG	4080
	ATAAACCAAA	GCCAAATCCG	GATCCAAAAC	CAGATCCAGA	TAAACCAAAG	CCAAATCCGA	4140
	ATCCAAAACC	AGACCCTAAT	AAGCCAAATC	CTAACCCGTC	ACCAGATCCC	GATCAACCTG	4200
30	GGGATTCCAA	TCATTCTGGT	GGCTCGAAAA	ATGGGGGGAC	ATGGAACCCA	AATGCTTCAG	4260
	ATGGATCTAA	TCAAGGTCAA	TGGCAACCAA	ATGGGAATCA	AGGAAACTCA	CAAAATCCTA	4320
	CTGGTAATGA	TTTTGTATCC	CAACGATTTT	TAGCCTTGGC	AAATGGGGCT	TACAAGTATA	4380
35	ATCCGTATAT	TTTAAATCAA	ATTAATAAGT	TGGGCAAAGA	TTATGGAGAA	GTTACTGATG	4440
	AAGACATTTA	TAATATTATT	CGAAAACAAa	ATTCAGCCGG	AAATGCATAT	TTAAATGGAT	4500
	TACAACAGCA	ATCGAATTAC	TTTAGATTCC	aATATTTCAA	TCCATTGAAA	TCAGAAAGGT	4560
40	ACTATCGTAA	TTTAGATGAA	CAAGTACTCG	CATTAATTAC	TGGTGAAATT	GGATCAATGC	4620
	CAGATTTGAA	AAAGCCCGAA	GATAAGCCGG	ATTCAAAACA	ACGCTCATTT	GAACCGCATG	4680
45	AAAAAGACGA	TTTTACAGTA	GTTAAAAAAC	AAGAAGATAA	TAAGAAAAGT	GCGTCAACTG	4740
	CATATAGTAA	AAGTTGGCTA	GCAATTGTAT	GTTCTATGAT	GGTGGTATTT	TCAATCATGC	4800
	TATTCTTATT	TGTAAAGCGA	AATAAAAAGA	AAAATAAAAA	CGAATCACAG	CGACGATAAT	4860
50	CCGTGTGTGA	TTCTTTTTTT	TTATTATGGA	ATAAAAATGT	GATATATAAA	ATTGCTTGT	4920
	TCCGTGGCTT	TTTTCAAAGC	CTCAGGATTA	AGTAATTGGA	ATATAACGAC	AAATCCGTTT	4980
55	TGTAACATAT	GGATAATAAT	TGGAACAGCA	AGCCGTTTTG	TCCAAACATA	TGCTAATGAA	5040

AATATTAATG AACTTACTGT TGTAGCAATA ATAAATGCCA CGATACGATT ACCTTTAATC 5160
 GCATTAAATA ATTCTCCAAA GATTACTTTT CTGAATACAT ATTCTTCTAA TAAAGGACCA 5220
 5 ATAATAGATA CAAAGAAGAT AAATATAGGT ATTTTTCGAG CAATAATAAT TAGCTTTTCT 5280
 GTATTAGGAC TTA CTGTGTTG TCCACCATAA ATTTGCGTTA ATACAATGCT CACTACCATT 5340
 10 TGATAAATCA TTACCAATGC AAATCCAAGC AATGCCCATG GAATGATATA TTTTTTAGGT 5400
 TCTTTAACTT CTAATTCTAA TTTTGTGGA TTTTAAATTT TTAAATTAAT TAAAATAATC 5460
 GTCGTGGCGG CGATTAAAAA TAGAACAAGT TGTATGTAA TGA CTGCTTT AGTCAGTTCT 5520
 15 ATGCCACTAT ATTGTACAAA TGGTAATTTT TTTACAATGA GAAGCGGTAA AAATTGAGAC 5580
 AATATATAAA TAATAACAGT TAGCAATGAT GCCCATAATC tTGTCATAAT TTTCTCCAA 5640
 ATATTTGTTT ATAATTTATT TTATCGTAAA TAACTTGAAG TTACAAA CT TAATTAAAAG 5700
 20 GTTATGACTT GAAATTTTGA CCAATTTGA TTATTATAAA TGTATGTTAG CACTCTTTAA 5760
 TGTTAAGTGC TAACTTTTAG GTTTTTTAAG GAGGAACAAT CATGCTAAAA CCAATTGGAA 5820
 ATCGTGTGAT TATTGAGAAA AAAGAACAAG AACAAACAAC TAAAAGTGGn ATTGTTTAAC 5880
 25 TGATAGTGCT AAAGA 5895

(2) INFORMATION FOR SEQ ID NO: 2:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6796 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TTTGAAAAA CAAGGTACGA TTGGTTTAAAT AACATATATG AGAACCGATT CTACACGFAT 60
 40 TTCaGATACT GCCAAAGTTG AAGCAAAACA GTATATAACT GATAAATACG GTGAATCTTA 120
 CACTTCTAAA CGTAAAGCAT CAGGGAAACA AGGTGACCaa GATGCCCATG AGGCTATTAG 180
 45 ACCTTCAAGT ACTATGCGTA CGCCAGATGA TATGAAGTCA TTTTGTGACGA AAGACCAATA 240
 CCGATTATAC AAATTAATTT GGGAACGATT TGTGCTAGT CAAATGGCTC CAGCAATACT 300
 TGATACAGTC TCATTAGACA TAACACAAGG TGACATTAAA TTTAGAGCGA ATGGTCAAAC 360
 50 AATCAAGTTT AAAGGATTTA TGACACTTTA TGTAGAACT AAAGATGATA GTGATAGCGA 420
 AAAGGAAAAT AAAGTGCTA AATTAGAGCA AGGTGATAAA GTCACAGCAA CTCAAATTGA 480
 ACCAGCTCAA CACTATACAC AACCACCTCC AAGATATACT GAGGCGAGAT TAGTAAAAAC 540

55

	AAAGCGTAAC TATGTCAAAT TAGAAAGTAA GCGTTTTGTT CCTACTGAGT TGGGAGAAAT	660
	AGTTCATGAA CAAGTGAAAG AATACTTCCC AGAGATTATT GATGTGGAAT TCACAGTGAA	720
5	TATGGAAACG TTACTIONGATA AGATTGCAGA AGGCGACATT ACATGGAGGA AAGTAATCGA	780
	CGGTTTCTTT AGTAGCTTTA AACAAGATGT TGAACGTGCT GAAGAAGAGA TGGAAAAGAT	840
10	TGAAATCAAA GATGAGCCAG CCGGTGAAGA CTGTGAAATT TGTGGTTCTC CTATGGTTAT	900
	AAAAATGGGA CGCTATGGTA AGTTCATGGC TTGCTCAAAC TTCCCGGATT GTCGTAATAC	960
	AAAAGCGATA GTTAAGTCTA TTGGTGTTAA ATGTCCAAA TGTAATGaTG GTGACGTCGT	1020
15	AGAAAGAAAA TCTAAAAAGA ATCGTGTCTT TTATGGATGT TCGAAATATC CTGAATGCGA	1080
	CTTTATCTCT TGGGATAAGC CGATTGGAAG AGATTGTCCA AAATGTAACC AATATCTTGT	1140
	TGAAATATAA AAAGGCAAGA CAACACAAGT AATATGTTCA AATTGCGATT ATAAAGAGGC	1200
20	AGCGCAGAAA TAATATTTTT ATTCCTAGA TACATTTTAA GATTGTTAAA TAGAATCATT	1260
	AGTGAATCTT ATTTTAAAGA TAGTAAAGGA TTAATCTAAA TAAGTGCGGA TAATATAAAC	1320
25	ATAACAACAT AATTAAMAGA CATAAATGAC aATAAAAGGA GTATAGAAAT GACTCAAAC	1380
	GTAAATGTAA TAGGTGCTGG TCTTGCCGGT TCAGAAGCGG CATATCAATT AGCTGAAAGA	1440
	GGAATTAAAG TTAATCTAAT AGAGATGAGA CCTGTAAAC AAACACCAGC GCACCATACT	1500
30	GATAAATTG CGGAACCTGT ATGTTCCAAT TCATTACGCG GAAATGCTTT AACTAATGGT	1560
	GTGGGTGTTT TAAAAGAAGA AATGAGAAGA TTGAATTCTA TAATTATTGA AGCGGCTGAT	1620
	AAGGCACGAG TTCCAGCTGG TGGTGCAATTA GCAGTTGATA GACACGATTT TTCAGGTTAT	1680
35	ATTACTGAAA CACTTAAAAA TCATGAAAAT ATCACAGTTA TTAATGAAGA AATTAATGCC	1740
	ATTCCAGATG GATACACAAT TATCGCAACA GGACCACTTA CTACAGAAAC CCTTGCGCAA	1800
	GAAATAGTGG ACATTACTGG TAAAGATCAA CTTTATTTCT ATGATGCGGC TGCTCCAATT	1860
40	ATTGAAAAAG AATCTATTGA TATGGATAAA GTTTACTTAA AGTCCCGTTA TGATAAAGGT	1920
	GAAGCTGCAT ATTTAAACTG TCCTATGACT GAGGATGAAT TTAATCGCTT TTATGATGCA	1980
45	GTATTAGAAG CTGAAGTTGC GCCTGTAAAT TCATTTGAAA AAGAAAAATA TTTCGAGGGT	2040
	TGTATGCCTT TTGAAGTAAT GGCAGAACGC GGACGCAAGA CATTACTATT TGGACCAATG	2100
	AAACCAGTAG GATTAGAAGA TCCAAAGACT GGGAAACGTC CTTATGCGGT GGTTC AATTA	2160
50	AGACAAGATG ACGCTGCTGG TACACTCTAC AATATTGTTG GCTTCCAAAC GCATTTAAAA	2220
	TGGGGAGCTC AAAAAGAAGT CATTAAATTA ATTCCAGGCT TAGAAAATGT TGATATTGTT	2280
55	AGATATGGTG TGATGCATAG AAATACCTTC ATTAATTCAC CGGACGTATT AAACGAGAAA	2340

	TATGTAGAAA GCGCagcTAG CGGCTTAGTT GCAGGTATCA ATCTTGCGCA TAAAATATTA	2460
	GGCAAGGGTG AGGTAGTATT TCCGAGAGAA ACAATGATTG GAAGTATGGC TTACTATATT	2520
5	TCTCATGCTA AAAACAATAA GAATTTCCAA CCTATGAATG CTAACCTCGG GTTATTACCA	2580
	TCTTTAGAAA CTAGAATTAA AGATAAAAAA GAACGCTATG AAGCACAAGC TAATAGAGCT	2640
10	TTGGATTACT TAGAAAATTT CAAAAAACT TTATAAAATA GTTAGAAAGA CTAGATATGC	2700
	TATTCATTCT TAAGTCATCA ACGAGTAAGT AATGACTTTC TAAATGGAAA ATACTTATCC	2760
	TAGTCTTTTT AATTTTGGAA TTGTTACGTA TTTCTGACAA TTTAGAATTC GCATTCAAAA	2820
15	AATATCTAAA TAAATAACAC GCAATAAGTT GATTGATGTA ACATGTAAGA GAATGTTTTA	2880
	AATAAACTTT ATTTAAAAGG CAATGAAATA ATAAATGGCA AGGCTATTAA TAAAGACTTT	2940
	TAGTAATTAA TTTAAAAAAG AGGTATTCTA ATTAACAGGT TTTCCGATTA GTTACAATTA	3000
20	TTTAATTCTC AAAAGATTTA GAATTGATTA TCAAATTACT GTAAGCCCTT TGCTGTATAT	3060
	GCTACAATTC TTATTGATGG AGGGTAAATG TATTGAATCA TATTCAAGAT GCGTTTTTAA	3120
25	ATACATTGAA AGTTGAACGG AATTTTTCGG AACACACATT GAAATCATAT CAAGATGACT	3180
	TAATTCAGTT TAATCAATTT TTAGAACAAG AACATTTAGA GTTGAATACT TTTGAATACA	3240
	GAGATGCTAG AAATTATTTG AGCTATTTAT ATTCAAATCA TTTGAAAAGA ACATCTGTTT	3300
30	CTCGTAAAAT CTCAACGTTA AGAACTTTCT ATGAATATTG GATGACGCTT GATGAGAACA	3360
	TTATTAATCC ATTTGTTCAA TTAGTACATC CGAAAAAAGA AAAATATCTT CCGCAATTCT	3420
	TTTACGAAGA AGAAATGGAA GCGTTATTCA AAACCTGTAGA AGAGGACACT TCAAAAAAATT	3480
35	TACGGGATCG AGTTATTCTT GAATTGTTGT ATGCTACAGG CATCCGTGTT TCGGAATTAG	3540
	TAAATATTAA AAAACAAGAT ATAGATTTTT ACGCGAATGG TGTTACCGTA TTAGGAAAAG	3600
40	GGAGCAAAGA GCGCTTTGTA CCGTTTGGTG CTTATTGTAG ACAAAGCATC GAAAATTATT	3660
	TAGAACATTT CAAACCAATT CAGTCATGCA ATCATGATTT TCTTATTGTA AATATGAAGG	3720
	GTGAAGCAAT CACTGAACGC GGTGTACGAT ATGTTTTAAA TGATATTGTT AAACGAACAG	3780
45	CAGGCGTAAG TGaGATTCAT CCCACAAGC TCAGACATAC ATTTGCAACG CATTTATTGA	3840
	ATCAAGGTGC AGACCTAAGA ACAGTACAAT CGTTATTAGG TCATGTTAAT TTGTCAACAA	3900
	CTGGTAAATA TACACACGTA TCTAACCAAC AATTAAGAAA AGTGTATCTA AATGCACATC	3960
50	CTCGAGCGAA AAAGGAGAAT GAAACATGAG TAATACAACA TTACATGCAA CAACAATTTA	4020
	TGCTGTAAGA CATAATGGGA AAGCAGCTAT GGCTGGAGAT GGGCAAGTAA CGCTTGGTCA	4080
55	ACAAGTCATC ATGAAACAAA CGGCAAGAAA AGTGCGACGT TTATATGAAG GTAAAGTGTT	4140

	ATTACAACAG TTTAGTGGTA ACTTAGAAAG AGCTGCTGTT GAATTGGCAC AAGAATGGCG	4260
	AGGCGATAAA CAATTACGTC AATTAGAAGC TATGCTAATT GTAATGGATA AAGATGCTAT	4320
5	TTTAGTTGTC AGTGGAACTG GCGAAGTTAT TGCTCCAGAT GATGACCTTA TCGCTATTGG	4380
	ATCAGGAGGC AACTACGCAT TAAGCGCAGG ACGTGCATTG AAACGCCATG CATCGCATTT	4440
10	GTCTGCTGAA GAAATGGCAT ATGAGAGCTT GAAAGTAGCG GCTGATATT GTGTCTTTAC	4500
	CAACGATAAT ATTGTTGTG AAACACTATA ATAATCAGAG CACGATAAAT AATTACGAGC	4560
	AATTAATTTT AGTTAAAAGA CGGAGGAATG AAATTAATGG ATACAGCTGG AATAAGATTA	4620
15	ACTCCAAAAG AAATCGTATC TAAATTAAAT GAATACATCG TTGGACAAAA TGATGCTAAA	4680
	CGTAAAGTGG CAATTGCCCT ACGTAATCGA TACAGAAGAA GTTTATTAGA TGAGGAATCA	4740
	AAGCAAGAAA TTTCACCTAA AAATATTTTG ATGATTGGAC CAACCGGCGT TGGTAAAACT	4800
20	GAAATTGCAA GAAGAATGGC CAAAGTTGTC GCGCGCCAT TTATAAAAGT AGAAGCTACT	4860
	AAATTTACTG AGGTAGGTTA TGTAGGACGA GATGTTGAAA GTATGGTTAG AGATCTTGTT	4920
25	GATGTTTCAG TAAGATTAGT CAAGGCGCAG AAAAAATCAT TGGTACAAGA TGAAGCAACA	4980
	GCTAAGGCCA ATGAAAAACT TGTTAAGTTA TTAGTTCCAA GTATGAAAAA GAAAGCGTCT	5040
	CAAACGAATA ATCCTTTAGA GTCACTTTTT GGAGGTGCAA TTCCAAATTT CGGACAAAAT	5100
30	AACGAAGATG AAGAAGAACC ACCTACTGAG GAAATTAAAA CAAAACGTC TGAAATTAAG	5160
	AGACAGCTAG AAGAAGGCAA ACTTGAAAAA GAAAAGGTAA GAATTAAAGT CGAACAAGAT	5220
	CCTGGTGCTT TAGGTATGCT AGGTACAAAT CAAAATCAGC AAATGCAAGA GATGATGAAT	5280
35	CAATTAATGC CTAAAAAGAA AGTTGAGCGA GAAGTTGCTG TTGAGACGGC AAGGAAAATC	5340
	TTAGCTGATA GTTATGCGGA TGAAC TAATT GATCAAGAAA GCGCTAACCA AGAAGCGCTT	5400
40	GAATTAGCAG AACAAATGGG TATCATCTTT ATAGATGAAA TCGACAAAGT TGCGACGAAT	5460
	AATCATAATA GTGGTCAAGA TGTCTCAAGA CAAGGTGTTT AAAGAGATAT TTTACCTATA	5520
	CTTGAAGGTA GCGTTATTCA AACCAAATAT GGTACTGTGA ATACTGAACA TATGCTGTTT	5580
45	ATAGGTGCTG GAGCTTTCCA TGTATCTAAG CCGAGTGACT TGATACCAGA ATTGCAAGGT	5640
	CGTTTTCCGA TTAGAGTTGA ACTTGATAGT TTATCGGTAG AAGATTTTGT AAGAATTTTG	5700
	ACAGAACCAA AATTGTCATT AATTAAACAA TATGAAGCAT TGCTTCAAAC AGAAGAAGTT	5760
50	ACTGTAACT TTACCGATGA AGCAATTACT CGCTTAGCTG AGATTGCTTA TCAAGTAAAT	5820
	CAAGATACAG ACAACATTGG TGCACGTCGA CTTCATACAA TTTTAGAAAA GATGCTAGAA	5880
55	GATTTATCAT TCGAAGCACC AAGTATGCCG AATGCAGTTG TAGATATTAC CCCACAATAT	5940

AAATATACAA AAGGAGAAAA ATTCATGAGC TTATTATCTA AAACGAGAGA GTTAAACACG 6060
 TTA⁻CTTCAAA AACACAAAGG TATTGCGGTT GATTTTAAAG ATGTAGCACA AACGATTAGT 6120
 5 AGCGTAACTG TAACAAATGT ATTTATTGTA TCGCGTCGAG GTAAAAATTTT AGGATCGAGT 6180
 CTAAATGAAT TATTAAAAAG TCAAAGAATT ATTCAAATGT TGGAAAGAAAG ACATATTCCA 6240
 10 AGTGAATATA CAGAACGATT AATGGAAGTT AAACAAACAG AATCAAATAT TGATATCGAC 6300
 AATGTATTAA CAGTATTCCC ACCTGAAAAC AGAGAATTAT TCATAGATAG TCGTACAACT 6360
 ATCTTCCCAA TTTTAGGTGG AGGGGAAAGA TTAGGTACAT TAGTACTTGG TCnAGTACAT 6420
 15 GATGATTTTA ATGaAAATGA TTTGGTACTA GGTGAATATG CTGCTACAGT TATTGGTATG 6480
 GAAaTCTTAC GTGAGAAGCA TAGTGAAGTA GAAAnAGAAG CGCGCGATAA AGCTGCTATT 6540
 ACAATGGCAA TTAATTCATT ATCTTATTCT GAAAAAGAAG CGATTGAACA TATCTTTGAA 6600
 20 GAAC⁻TGGCG GTACGGAAGG CCTATTAATC GCATCAAAAG TTGCAGATAG AGTTGGTATT 6660
 ACTAGATCTG TAATTGTAA TGCCTACGT AAATTAGAAA GTGCTGGTGT AATTGAATCA 6720
 CGTTCTTTAG GAATGAAAGG TACTTTCATT AAAGTTAAAA AAGAAAAATT CTTAGATGAA 6780
 25 TTAGAAAAAA GTAAAT 6796

(2) INFORMATION FOR SEQ ID NO: 3:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2073 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCC⁻TAAAAAT TnAAAATTAT CACGCCTTTT GaACAGCTTT GTAACCaTCt GGACGATCAT 60
 40 kAAATTC⁻CaA TGTAATCCTT GGTTTAAaGT TGATCTTTAA CCTTATTTAA AycACCAATT 120
 GTACGTATAT TATGTTGTTT AGCAAAATCA CGTTTACAG CTAAAGCATA CGTATTGTTA 180
 45 TACTTCATTG GTTTTAACAT AGTCATTTGA TATTTCTTTT CAAGACTTTG CTTAGCTTGT 240
 TCATAAACTT TTTTCTCTTC TTTGACTTC AATGGTCTT TTGTTAATTC ACCTAAACT 300
 GTTCCAGTAA ATTCTAAATA CCCATCTATA TCGTCAGATT TTAAAGCATT AAATAAAAT 360
 50 GCTGTTTTGC CCATACCATC TTCACTTCT ACAGATTTT TGGTCTCTTC TTCTATTAAA 420
 ATTTTATACA TATTTGTAAT AATCGATGGC TCGGAGCCAA GCTTTCCAGC TAACGTAATT 480
 TTATCACCTT TTTGTGCAAA CATAGGAATA GCGATAGCCA GTATAATAAT CATCACTATA 540

55

TCAAATATAA TTGCCAATAA GGCTGCTGGA ATTGCACCTA ATAATATCAA CGATGCATTG 660
 TTACGGTCTA TACCTAATAA AATTAAATCT CCTAGTCCGC CTGCACCAAT TAATGCTGCT 720
 5 AGTGTTGCTG TACCTATAAT TAATACCATA GCCGTTCCTA CACCAGCCAT TATAACAGGC 780
 ATTGCTATCG GAAGTTCGAC TTTAGTTAAA CGTCTAAATG GTTTCATACC TATACCTTTA 840
 GCCGCTTCAA TGAGTGATGG ATCAACTTCT TTAATTCCAG TATACGTATT CCTTAAAATT 900
 10 GGTAACAACG CATACTACTAC AAGTGCAATA ATTGCTGGCA CACGACCGAT ACCAAATAAA 960
 GGAATCATTAA AACCTAATAA TGCCAACGAT GGTATGGTTT GAAGAATTGC CGCAATATTC 1020
 15 ATTACGATTT CAGATATCGT TTTAGTCTTC GTTAATAAAA TACCTAATGG TACCGCAATA 1080
 GCAGTTGCAA TCAATAATGC GATAAATGAT ATTTGAATAT GTTCTATCAT TGTCGAAAAG 1140
 AGTTGCCCCT TACGTTCACT CAATATGTCg AAAAAAGTTAG TCATGTTGAG CTACCTCCTT 1200
 20 TTTCTGGGAC AAATATTTGA AGATATCTTT CCTATCAATA ACATATTGAC CTACGCTATC 1260
 TTCTTGCAATG ACAATGACAC GCTCGCTCTC TGATAAAAGT TGATACAATA CTTCAATTGG 1320
 TTGATTGTCA TAAACAATTG GATAAGCGCT CATAGATGTA ACCTCATCGA TTGGTTTCAT 1380
 25 AATATCCAAG TCACGGATAA TTGCGTTCTC TTCAACACAT GGCGCATCAT CTTCTAAATG 1440
 ACTACCCATA AATTGTTTAA CAAATTCCTT TTGAGGATTA TTTTAAATC CTTCTGGTGT 1500
 30 GTCAATTTGT TCAATATGCC CTTCAATCAA AAGACAAATC TTATCACCAA GTTTCATCGC 1560
 CTCTTGAATA TCATGTGTAA CAAATATGAT TGTCTTCTTA ATTTTAGTTT GTAATTCAAT 1620
 TAAATCATCT TGAAGTTTTT CTCGGCTGAT TGGGTCTAAT GCACTAAACG GTTCATCCAT 1680
 35 TAAAATAACT GGTGGATCAG CTGCTAACGC ACGTATAACT CCTACACGTT GTCGTTGCCC 1740
 CCCTGACAAT TCATCAGGTT TTCTGTTTTT ATATTTTTCA GGTTCTAATC CAACCATTTT 1800
 AAGTAATTCA TCTACTCTTT TATCTATATC TTTTCTTTT CACTTTTTCA TTTGTGGCAC 1860
 40 TTGTGCAATa TTTTCTTTGa wTGTCaTATG TGGGAATAAT GCAATCTGCT GCAATACGTA 1920
 TCCAATATCC CAACKCATTT CGTATACTGG ATAATCACTT ATTGGTTTAT CTTTAAAATA 1980
 45 AATATAACCT TCACTTAAGT GAATGAGTCG ATTAATCATT TTTAATGTCG TAGTTTTTCC 2040
 ACAACCTGAA GGTCCAATTA GCACAAAAAA TTC 2073

(2) INFORMATION FOR SEQ ID NO: 4:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13321 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

	ACTATTCTAG CTTTCATCAGT TATCATATAT TCTTTGAAAC ACTTGTAAGA AAATATAATG	60
5	AGTATTTACT ACATAATGAT ATTTCAAATT AGAAAAAAGG AAGTTATGAT TTAATGGCCT	120
	TGAGCCTATC ATAACCTCCT TTTATCATT TATTGTTGTG TTGATGTTTC GATAACGTGG	180
10	TACATCTTAT CAAACATCAA TTCGAAACCA TGCACCATGG CATCATGATA TTCTTTTTTC	240
	TTTTGCTTGT ATTCTAAATT AGTAAATCGT CTTTCTTTT CAACTAATGA ACGATAATAA	300
	AATAGCATTT GGGTGCCACC TGTTTCACGT TCAAAAAATT CTACCTCAAT GACATCTTGC	360
15	GTTTCACTTA GTCCAGGCAT ACCGATAGTC ATCTTAACGT ATTCATCCAT AACTAAAGAT	420
	TCATAAATGC CTTCAATCAC ATTTACTTTG CCATTACGTT GTTGATCTAC AATACGATAT	480
	TTACCGCCTT CTTTAACGTC CGCTTCAATC TCTTTATTCG TTCTGGCTGA TGTCAATAAC	540
20	CATTGTTTCA ACAAATCTTT CTTTGTCCAA GCTTCGTATA CTAACCTCTGG AGAAAAATTA	600
	TAAAGCTTTT CAATTTCAAC TTCGACATGT TCATTCTCTA CATTAAATTT TGCCACTGTT	660
	GTCCACCCAC TTTGCTCTT ACTTTTATTT TAACGTATTT TTGCTCAGTT CCAAACATAG	720
25	ATGATCATCA TTTTAAAG ATTAGCGTTA TACGGTGAGT ACAACATGAT CTGTTAATAT	780
	AACAAGCCAC CTTACTTGGC TACATCGATA TATTGTTAAG CATTAAATGTT TCATTTCTTG	840
30	ACTAGTGTTT TTTTATAGCT TTGAAAATT AAATAAAATC GCAATAAGTC CGCATACACC	900
	TAATAATATA GGATAAATGC TGTATGGGAA TAACATTAAC GGTGAAATAC CAGCTACACC	960
	AGCCGCTGaa ATGACTTGCG GGCTATATGG TAATAAACCT TGGAAGCAGC CTCCAAATAT	1020
35	ATCAAGAATA CTTGCTGATT TCCTTGAATC TACATCATAT TCATCTGCAA TATTTTATAGC	1080
	TAAAGGACCT GACATAATAA TAGAGATGGT GTTGTGTTGCC GTGGCAATAT CTGCGACACT	1140
	TACCAAACTA GCAATTCCTA ATTCTGCGCC ACGCTTTGAT TTCACTTTAG AGCGAACAAA	1200
40	TTGCAACAAC CATTCAATAC CACCATTGTG TTGAATAATA CCGACTAAAC CACCAATTAG	1260
	CAACGCAATC ATAGCAATAT CTTCCATGCT TATAATACCT TTGGACACTG CATCTAGTAG	1320
45	CCCCATCCAA CCGAATGAAC CATCTATGAG ACCAATGATT CCGGCTAATA ATGTTCCGCC	1380
	AATCAATACG ATAATGACAT TTACACCTAA TAATGCTAAT ACCAATACTA AGATATACGG	1440
	TACAACTTTA ATTAGATTAT AATCATAGTt TTTAGCATGA TTTAAAGAAA TGCCATTTCGT	1500
50	TAAGAAATAC AGAATAATAA TCGTTAAAT AGCACCTGGC AATACAATTT TAAAGTTTAC	1560
	TCTGAATTTA TCTTTCATTT TCGTATGTTG TGTCTAACC GCAGCAATTG TTGTATCTGA	1620
55	AATCATTGAT AGATTATCGC CGAACATTGC ACCTCCAACA ACTGTAGCCa tTGctAGCGC	1680

	TCCTACAGAC GTCCCATAG ATATAGAAAC AACATACAA ATCACAAACA ATCCTACAAT	1800
	AATTAAATTT TCTGGGATTA ATGATAGTCC TAAATTAAGT GTCGACTTTA CGCCACCCAT	1860
5	TTTTTCAGCT GTATTTGAAA ATGCACCTGC TAAAATAAAA ATCAACATCA TTAAAACAAT	1920
	GTTTGAATGG CCTGCACCTT TCGTGAAGAC CTCAACTTTT TTAGCAAATG ATTCTTTTCG	1980
10	ATTCATTAAT AACGCCACAA TTACCGTTAT CGTAATTGCA ACATTTAATG GCATTGAAGT	2040
	AAAATCACCT GTGATAATAC CTACGCCTAA AAACAACGCC ACAAATAATA ACAAGGGGAA	2100
	TAATGCCCAA GCATTGCTCT TTTTATGTAC TTCCATCCTT TTTACCTGCT TTCCAATTAA	2160
15	AAATACCTCT TTCTCACAAA CGATGAAGAA AGAGGTTTTC ATGTGCTTTA CCTGCTTATC	2220
	TTCAAACCAT TACGGTTACT GGAATTGGCA CAFTCGAGAT GTTGCCGAGG CTTCATAGGG	2280
	CCAGTCCCTC CACCTCTCTA GATAAGTGAT GCTTATTTAC GTTTACGTTA CAAGATAATC	2340
20	CTTAGTACGT CAATCATAAA TTAATCAGGA GTCGTATAAT ATTTTTCATA AACATCATT	2400
	GCTACTGTAA TAATAATCAA AACATAATG CTAATAACAA GTAAAAGCCA CCATTTAAGC	2460
25	ATTAATGCAA TAAAAATGAA CACGATAGAC ACACCTACTA ATATTAATGA TATGACTTTA	2520
	AATTGCTGAA CACGTTGCTT GGAGATGACT TTCAACTGTT TGTTTGATAG ACGCGTATTT	2580
	TTTATACTGA TTCCAGTAT ATTTTCTAAT ATTTGAACCA ATACGATACT TATTGCAAAT	2640
30	ATAATAATTG GTAAAACATC ATAGCTCCCT ATAGTTAATG TATAAATTAC AAATCCAATG	2700
	TAAAGTAACC CTGAGACAAA GGATAAAAAG TATGCGACGT ATTTGTTAAA CTTAATGATA	2760
	TGCTTTTTTAA CGTTTTGATG TGTAACCAT ACATTCGAAA CGATCGCAAC TGCTACAAAT	2820
35	AATGTGAATA CTATATATAA TGGTAATTTT TGFTCAGGAA AAACAGTCGC TATTCCAAAA	2880
	GCTAATGCTA AAATCAAAAA TAATATAGCT CTAGATACTA TTAATGCCAT AATAACAACC	2940
40	CCTTTGTTTA ATATCGAGTT TGCAAATTTA CGTTTATCAG CGTTTCTATG ATCAGTACTT	3000
	CTACGGGTAG CGTTTCTATG TAATTTACAT CATCTTAACA TATAAATACT TCGCTATTTA	3060
	ATTGAAAACA TATCCTATTA TTCTTTGTCC GTTCTGACGT TTAATATCTA GCCTTAGGCA	3120
45	TTTCACTTGT TAATGAATTT AACTTTCTTC CACTAACCGT CCCTAAACCC AATCCCGCAA	3180
	CAGTTTTTAA CTTTTTCGTT GTTGTCTGTA CATCCTCATT AAGAAAGTTT ATTCTGCTTA	3240
	AAACTTATAA TCCACACCCT GAGCAAACGC TCCTTATGAC AGAGTATTAA AATAAGCCGA	3300
50	TAAAGATACA CACCTTTACC GACTATTTAA AATACACTTC ACCAATTCAT TTTAATTTAA	3360
	TGGATTGAAG TAACTAAATT AATATTATGT TGTTCAATTA AAAGCTTCAT ACAAACCTAA	3420
55	TCTATTTGCA CTCCACCGCT AACACCGAAC ACTTGTCGGT TTGTATAACT TGATTCTTCT	3480

	GTGTTTGGAC CAAATGTTGG GATTTTACTT TGAGGTTGTC CACCAGAAAT TTGTAATGGT	3600
	GACCAGAATG GACCAGGCGC TACACAGTTC ACTCTAATTC CTTTGGTCC TAATCTTCT	3660
5	GAAAACTTT TAGTTAATGA AATAATTGCT GCTTTTGAAG CGGCATAATC ATGAAGAATA	3720
	GGACTAGGAT TATAACCTTG TACAGATGAT GTCGTTGTAA TTGACGCACC CGGTTTTAAA	3780
	TATCCAATG CTTTTTGAAC TGTCCAAAAT AGCGGATAGA CATTGTTTC AAATGTTTCT	3840
10	GTAAATGCCT CAGTTGTAA TCCATGAATA TCATCATGAT ACTGTTGATG TCCAGCAACT	3900
	AAAGTAACAT TATCTAAGCC ACCTAATTGT TGATATGCTT GTTCAACAAG GTCATAGTTG	3960
15	AACTGTTTCT CTCTTATATC ACCAGGAATT AACACTGCCT TTTGACCACT TTCTTCAATC	4020
	ACTTGCGTA CTTCTTGTGC ATCTTGTCT TCACTCGGAA GATAGTTAAT CGCTACATCT	4080
	GCACCTTCTT TAGCATACGC AATTGCTGCT GCACGCCCTA TTGCTGAGTC ACCACCTGTG	4140
20	ACTAATATTT TATAGCCTTG TAAGCGTTGA TGACCTGGT AAGACGTTTC GCCACAATCG	4200
	GGTGCTGGCG TCATTTTCTA TTGTAAACCC GGTACCTCTT GTTCTTGTTC TTCATAATCC	4260
	GTTGTTTTAA ATTTTGTCT AGGATCTTGA GCTGCCATTT TTTTACATCT CCTTATTCGC	4320
25	TTAATGGTTA TTATTTACCC AATCTTCCTA GGAACCTAAT CATGATTACA CTAAAAATTA	4380
	CTTCTTCTT TATAAAAACA AGCTCGAATT ATTCATGCAA TAGTCTCTTT ACAAATTCAA	4440
30	CAAAATACTC AGGTACTTTT TCCAGAATCC TTTCATCCGG TTTATATTGA GGATGATGTA	4500
	AATCATATTC ACTATGAGAA CCAATTAACG CAAATACACT TGGAAAATGT TGAATATAAC	4560
	CTGAAAAATC TTCTCCAATC GTAAGCGGCT GTTCCATCAT TCCCACCTTA TATCCAACAT	4620
35	GTTGGGCTAC TGCAATTGCT TTATGCGTCA ATGCCTCATC ATTCATCACA GCGCCAGGTA	4680
	AATGCGTATA ATTTAAATTA ATTTTCATAT TATATGCTTG AGCCAATCCG TCCGCAATAT	4740
	CTTGTAATCG TGTTTCTACA AGCTTTTCGTA CCACAGGATC AAAACTACGC ACTGTGCCTT	4800
40	GTACATACGC ATGATCAGCA ATGACATTCC AAGTATTACC ACATGATATT TGTCCAATTG	4860
	TTACTACCGC TTCATCAAAC GCAGATAGAT TTCTACTAAC TATGGATTGA ATACTATTAA	4920
45	TCAATTGCGC CAACACAATA ACTGGATCGT TGCATTGTTC TGGCTTTGCA GCATGACCAC	4980
	CCACGCCTTT AATATGAAAC TCAAAACGAT CTACTGCTGA TGTAATTGCC CCTGTTTTGA	5040
	TTGCAAATGT ACCTACCGAA CGCGATGGGT CATTATGAAA ACCCAATACT GCTTGTACAT	5100
50	CTTTTAATGC ATGTGTTTCA ATAATTTTAA AAGCGCCATG TCCTAGTTCT TCTGCTGATT	5160
	GAAAAATGAA TTTAACACGC CCAGTAAGAG TGCCCTCAAT TTCTTTTAAT TTTACAGCTG	5220
55	TAGCCAAAAT ACTAGCCATG TGAATATCAT GACCACACGC ATGCATAACA CCTTCATTTT	5280

	CAGCTATACA	ACTCAGACCT	TGTCCCACTT	CAGCAACAAG	CCCAGTCGCA	AGTGGTAAGT	5400
	CTAATATTCT	AATATGATGT	TCTGTTAAAA	TATCTTTAAT	TTTTTGTA	GTCTTAAATT	5460
5	CTTTATCGGA	TAGTTCTGGA	AATTGATGAA	AATACCTTCT	CCAGGTAACA	GCTTGATCTT	5520
	TTAATCCCAT	CGGTCATTCC	CCTTCCTTAA	GTCAATGATA	TGTTGTCTAC	CCTACGATGA	5580
10	TCATCTTTGA	CTATTAAACG	ATGATTTTCA	AACAATGTAC	TCTTGTTAAT	TGCTTTCGTT	5640
	AATGATAGAC	AGTTGTTTAA	TAATATCGTA	ACACTGTTGT	CAAACATTC	TAACTTTTAT	5700
	AATTGAGACT	CTATACAAAA	ACGTGTTCTC	GAATATACTT	GTTTTTACAA	ACCACAAAAA	5760
15	GCTCTAAACA	TTAGTTTAAA	CCAATGCTTA	GAGCTTTCTA	ATTATTTTAT	GCTTTAAAAG	5820
	ATACTGTGTT	ATCTACGATG	ACCTTACCGT	CTTAAATAAC	TTTTTCTGCG	TGATTGATAC	5880
	CAAAATGATA	TGGAATATAT	TCATGATTTG	GTGCATCCCA	AATTACTAAA	TTAGCCTTAT	5940
20	CACCTGTGTT	AATTGTACCC	GCGTTAATGT	CTATTGCTTT	AGCAGCATTG	ACCGTAACAG	6000
	CATTCCAAAC	TTCATTAGGT	GATAGCTTTA	ATTTCAGGC	TGCAATCGCC	ATAACAAGTT	6060
25	GTAAGTTGTT	TGTGACACTA	CTACCAGGGT	TATAATCAGT	TGCTAATGCA	ATCGCACCGT	6120
	TATTGTCAAG	CATGCCTCTT	GCATCTGCAT	AATCTTCTTT	ACCTAAATAG	AACGTCGTTG	6180
	CAGGTAAGAG	GACAGCTACA	GTATCACTAT	TTCGCAACTT	TTCTTTTCCT	TTATCACTAG	6240
30	AAGCTACTAA	GTGGTCTGCT	GATATTGCTT	GTCATCAAT	TGCTAATTCC	AGTCCGCCTA	6300
	ACGGATCAAT	TTCATCCGCA	TGTATTTTCA	CTTTAAAACC	TGCTTCTTTG	GCTTTTTCGA	6360
	TATAATGTTG	CGATTGTTCT	ATTGTAAATA	CACCTGTTTC	ACAGAAAATA	TCCGCAAAGT	6420
35	CTGCATATTG	TTTACTTCC	GGAAGTAACG	CAATCATTTT	TTCTAAAAAT	GCCTCATTTG	6480
	AACTTGCCCTC	TTTAGGTACA	GCATGAGGCC	CTAGGAAAGT	ATGTTTCATG	TCTAAATCAT	6540
40	ATTTCTCAGC	TAAACGATTA	GACACTTTCA	ATTGCTTCAG	TTCATTTTCT	CTATCTAATC	6600
	CATAACCACT	CTTACTTTCA	ACTGCAAGCA	CGCCGTGTTT	AATCATAGTA	AGCAAATCAT	6660
	GCTCTGCTTT	TTTAAACAAG	TCATCTTCGG	ATGTTTCTCT	AGTAGCATT	ACGGTAGATA	6720
45	ATATGCCACC	ACCCATTTCT	AATATTTCAA	GGTAAGACTT	ACCTTGACGT	TTAATGACA	6780
	TCTCATGTTT	TCGAGATCCA	CCAAATGTTA	AATGGGTATG	TGCATCTACT	AATGCTGGGG	6840
	ACACTACCTT	CCCACTAGCA	TCAATCGTCT	CAGTCGCATC	GTAGTCATCT	GTATGTGTTC	6900
50	CAGCATATAC	AATTTTGCCA	TCTTTAATGA	CAACTGTACC	ATTTTTCACA	ACATTTAATT	6960
	CATCTAATTC	CTTACCCTTC	AAAGGTTTAT	CTGTTGATCT	CGGTAAATTC	AATTCTGCTA	7020
55	TATGATTAAT	TATTAAATCA	TTCATTACTT	ATCACCTGCT	TTATCAATCA	TTGGAATATG	7080

	AACACCCATA	CCTGGGTCAG	TCGTCAATAC	ACGTTCCAAT	CTTCTTTCAG	CACGCTCTGA	7200
	TCCATCTGCT	ACAACAACCA	TACCCGCATG	AAGTGAATAT	CCCATGCCAA	CACCGCCACC	7260
5	GTGATGGAAT	GAAATCCATG	AACCACCTGC	AGCTGTGTTA	ATGAGTGCAT	TCAATACAGC	7320
	CCAATCACCA	ACCGCGTCAC	TACCATCTTT	CATACTTTCT	GTTTCACGGT	TAGGACTAGC	7380
10	AACTGAACCA	GCATCTAAAT	GGTCTCGTCC	AATAACAATT	GGTGCTGAAA	TTTCACCGTC	7440
	ACGTACAAGA	CGATTTAAAG	CTAAGCCCAT	TTTCGCTCTT	TCTCCATAGC	CTAACCAAGC	7500
	AATACGTGAT	GGTAGTCCTT	GATATGAAAT	TTTTTCTTCA	GCTAAATCAA	GCCATCTTAA	7560
15	TAACTTTTCA	TTTTCTGGGA	AAAGTTTGCG	CATTTCTTCA	TCCGCACGCT	CGATATCTTT	7620
	TGGATCACCA	CTCAACGCAG	CAAAGCGGAA	TGGCCCTTTA	CCTTCACAGA	ATAATGGTCT	7680
	AATGTAAGCT	GGTACAAAGC	CTGGGAAGTC	AAAAGCATTT	TTCACCTCCG	TATTGAAGGC	7740
20	TACTTGACGA	ATATTGTTAC	CATAATCAAA	TGCTACAGCG	CCACGTTTTT	GGAATTCAAG	7800
	CATTAATTCA	ACATGCTTTG	CCATTGAAGC	TTGTGACAGT	TCAACATATT	TTTTCGGATC	7860
25	TTTTTCACGC	AATACTTTTCG	CTTCTTCTAC	AGAGTATCCT	TGTGGCACAT	ATCCATTTAG	7920
	CGGATCATGT	GCACTTGTTT	GGTCAGTAAT	AATGTCAATT	TTAAATCCTT	TTTCTAGAAT	7980
	CGCTTGATGG	ATGTCTACAG	CATTTCCAAC	TAACCCGATT	GATAATCCTT	CTCCACGTTC	8040
30	TTTCGCCTCT	TCTGCTAATT	TTAATGCTTC	ATCTAAATCA	GCTGFTTTAA	CATCACAGTA	8100
	TTTCGTATCA	ATTCGCTTAT	CAACACGTGT	TTCATCAACA	TCCACGCAAA	TTGCTACCCC	8160
	ATGATTCATA	GTAATTGCTA	ACGGTTGCGC	ACCACCCATA	CCACCTAAAC	CTGCTGTCAG	8220
35	TGTAACAGTG	CCTGCTAAAT	CTCCATTAAA	GTGTTGATTA	CCTAGCTCGG	CAAATGTCTC	8280
	ATAAGTACCT	TGCACAATAC	CTTGAGAACC	AATATATATC	CAACTACCGG	CTGTCATCTG	8340
	TCCATACATG	ATTAAACCTT	TTTTATCTAA	TTCATTAAAA	TGATCCCAGT	TTGCCCATTTC	8400
40	AGGCACTAAT	ACTGAATTG	AAATTAATAC	ACGTGGCGCT	TCTTCATGTG	TTTTAAATAC	8460
	AGCAACTGGC	TTTCCTGATT	GTAATAACAT	TGTCTCATCT	GATTCTAATT	CTCGTAACGT	8520
45	TTTCTCTATT	GCTTCAAAAAG	CTTCCCAATT	ACGTGCTGCT	TTTCCAATAC	CACCATAAAC	8580
	AACTAAATCT	TCTGGTCTTT	CAGCAACTTC	TGGGTCTAAA	TTGTTGTATA	ACATTCTAAG	8640
	TACTGCTTCT	TGTTCCCAAC	CTTTACACTC	AATACTCAAA	CCTTTTTTTG	CTTGAATTTT	8700
50	TCTCATAAAA	TTGCTCCTG	TTCTTTTAAG	AAGTTAATTC	CACTAAATTT	AAAACGCTTA	8760
	CATTATTATC	TTCAATATTC	ATTATAGTAT	GTTAAAAATAT	AGCCAACAAA	TATAAAATAA	8820
55	CTAATTATCC	ATAGCTTGAA	TCTATAAATA	AAAGGAGCAA	AACACATGAA	AATTATTTCAG	8880

	CATATTAGCC	AGCCATCTTT	AACTGCTACG	ATTAAAAAAA	TGGAAGCAGA	TTTAGGTTAT	9000
5	GACTTATTTA	CACGTTCAAC	AAAAGACATC	AAGATTACCG	AAAAAGGAAT	ACAGTTTTTAT	9060
	CGTTATGCGA	GCGAATTAGT	TCAACAATAT	CGATCCACGA	TGGAAAAAAT	GTATGATTTA	9120
	AGCGTTACAT	CAGAACCAAG	GATAAAAATT	GGGACTCTTG	AATCTACGAA	TCAATGGATT	9180
10	GCGAATTTAA	TTCGAAAGCA	CCATTCCGAC	TACCCTGAAC	AGCAATATCG	TTTATATGAA	9240
	ATACATGATA	AACATCAATC	TATAGAGCAA	TTACTGAATT	TTAATATTCA	TTTAGCTATA	9300
	ACAAATGAAA	AAATAACCCA	CGAAGATATA	AGATCCATTG	CTTTATATGA	GGAATCTTAC	9360
15	ATTTTATTAG	CACCCAAGGA	AACATTTAAA	AATCAAAATT	GGGTAGATGT	TGAAAATTTG	9420
	CCACTCATAT	TACCAAACAA	AAATTCTCAA	GTGCGCAAAC	ACTTAGATGA	CTATTTTAAAT	9480
	AGAAGAAATA	TTCGTCCAAA	TGTCGTTGTA	GAAACAGATC	GATTGGAATC	AGCAGTTGGA	9540
20	TTTGTTTCATC	TCGGCTTAGG	TTACGCTATC	ATTCCGAGAT	TTTATTACCA	ATCATTTTAC	9600
	ACGTCTAATT	TAGAATATAA	AAAAATTCGT	CCAAACTTAG	GCCGAAAAAT	TTATATCAAT	9660
25	TACCATAAAA	AACGCAAACA	CTCCGAACAA	GTACATACAT	TCGTACAACA	ATGCCAAGAT	9720
	TATTTATATG	GACTTTTAGA	GGCTCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9780
	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9840
30	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9900
	CTCAGTCAAC	TGTATACCTT	TTTCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9960
	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGTGCCTCT	TATGTAGTTG	10020
35	CGTAGTCAaC	TGTaTACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	10080
	CGCAGATCAT	CGTATAAAAA	TTAATGACGT	CATTTCAAAA	ATCGATACAA	AAATAATTTA	10140
40	TTATAAAAAT	TCTAAGAAAG	AAGTGAAGCA	GATGTTAAAA	TCTATTAATC	ATATATGCTT	10200
	TTCAGTCAGA	AATTTAAACG	ATTCAATACA	TTTTTATAGA	GATATTTTAC	TTGGGAAATT	10260
	GCTATTGACT	GGTAAAAAAA	CTGCTTATTT	TGAGCTTGCA	GGCCTATGGA	TTGCTTTAAA	10320
45	TGAAGAAAAA	GATATACCAC	GTAATGAAAT	TCACTTTTCA	TATACACATA	TAGCTTTCAC	10380
	TATAGATGAC	AGCGAATTTA	AATATTGGCA	TCAGAGGTTA	AAAGATAATA	ACGTGAATAT	10440
	TTTAGAAGGA	AGAGTTAGAG	ATATTAGAGA	TAGACAATCA	ATTTACTTTA	CCGACCCTGA	10500
50	TGGTCATAAG	CTAGAATTAC	ATACTGGCAC	ACTTGAGAAC	AGATTAAATT	ATTATAAAGA	10560
	GGCTAAACCA	CATATGACAT	TTTACAAATA	AGGTGTCATT	ATAAAAAGGC	CTCTTGAAC	10620
55	CCGTTAAAAT	TTTAATTAAT	TATTATATAA	TAAGAGAACT	TTTCAAACAA	TACAGTTGTT	10680

	TTACTGCAAT TATTTTTCAA ATATATCAAC GTTAATATAA CTTCTATTAA GAAATACTCA	10800
	CATTCTGCCC TGCAATGCAA ATCTCGTCAC ATATAAATAT TTTTAATTAT TTTAAAAAAT	10860
5	GATGCACTAA ATTAGCAACG AGCTTAGCAG TTCTATTGTC AGCGTCATAT GTTGGATTCA	10920
	TCTCAGCAAT ACTAACTGAA GACACCTTAT CACTTGGAAT AATACGTTTT GCTAATTCAA	10980
10	GAACAGTATG TGGATACAAA CCTAACACTG CCGGCGCACT TACCCAGGC GCAAACGCAC	11040
	TATCAATGAC ATCCATACAA ATCGTAAACA TAATGACATC ATGTTTCATGT ACAAACGTT	11100
	CAATCATATC TTTAATTGTT GGTGATACGT GACTCAATAA TTCATCTGCA AAGACATAAT	11160
15	CAATCTTTTT CTCTTTAGCA TAATCAAATA AACTTTGCGT ATTACCACCT TGAGCAATAC	11220
	CAAGCACTAA ATAATCTGTG TTTTCATCTT CTTCTAAAAT TTGTCTAAAG CTCGTTCCAG	11280
	ATGTAGATTG TTGTTTCAGCA CGTGTATCAA AATGCGCATC AATATTTATC ACACCAATAG	11340
20	ATTGTGTTGG ATAGACTTTA CGTGTGCTA AATATTGAGC ATACGCAATA TCATGTCCAC	11400
	CACCTAATAA AAATGTTTGT CTATGATTAG CAATTGACTT CGTGCAAGC ATAGCAAATT	11460
25	CTTTTGTAGT ATCAATTAAT TCCTCATGAT CATGATAAAC ATTTCCGTAA TCGACTAAAG	11520
	TTcACATTGA TTCAAATCCG GCAAACCTGC AAATGCTTGT TTAATCGCAT CTGGTCCTTC	11580
	TTTTGCACCA ATGCGCCCCCT TGTTTAAAGC AACACCTTTG TCAACAGCAT AGCCTAATAT	11640
30	ACCGACCCCT GATGGCATAC TACTCTTTTC CAGCTTAGAC AAATCTTCAA ATGTTACTGT	11700
	TTGAAAATGT CTAAATTTTT TCGGGTCTGT TTCACTATCT AACCTTCCAG TCCATAAATT	11760
	TGGTTCACCT TGCTTGTA CAAGCATTTCC CCCTCTTATT TATGTGGCTT ATTAACAATT	11820
35	AAAGTATAAC GTATAGGAAA TTTTGAATTC AATTCATAGT TAAATCCGTA TCTTAAAAAT	11880
	ACTTATCTAC ATTACTTTTA CCCCTATTTT CTATGTAATA ACGAATACTT AGCTGATTTA	11940
40	TGTTAATAAA ATACGTCAAG ACTATTACAT TTTCATTAAT ATTGACATAG ACAATTTATC	12000
	TCTCGGCTTG TAATATGTAT AATTGTTACT AAAAGATATT TTGCTTGTTA CCTAATGGAG	12060
	GTTACATATA ATGAAGAACA ATAAAATTTT TGGTTTTCAA TGGGCAATGA CGATTTTCGT	12120
45	CTTCTTTGTC ATTACAATGG CGTTATCCAT TATGCTCAGA GATTTCCAGT CTATAATTGG	12180
	TGTCAAACAC TTTATATTTG AAGTTACAGA TCTAGACCA TTAATTGCTG CAATCATTTG	12240
	TATACTCGTT TTCAAATATA AAAAGGTCCA ACTTGCAGGT TTAATTTCT CAATCAGCCT	12300
50	GAAAGTAATT GAACGTCTAT TGCTAGCTTT AATTTTACCT TTAATTATTC TAATTATTGG	12360
	TATGTACAGC TTTAATACAT TTGCAGATAG CTTTATTTTA TTACAATCAA CAGGCTTATC	12420
55	AGTACCTATT ACACACATTC TGATTGGACA TATTCTGATG GCGTTCGTAG TAGAATTCCG	12480

	TGTTGTTGGT TTGATGTATT CAGTTTTCTC AGCAAATACA ACTTATGGTA CAGAATTTGC	12600
	TGCTTATAAC TTCCTTTATA CATTCTCATT CTCTATGATT CTTGGTGAAT TAATTAGAGC	12660
5	GACTAAAGGA CGTACAATTT ATATTGCAAC GACATTCCAT GCTTCAATGA CATTCCGACT	12720
	TATTTTCTTG TTTAGCGAAG AAATCGGCGA TCTATTTTCA ATCAAAGTCA TCGCCATTTT	12780
10	AACAGCAATC GTTGCAGTAG GATACATTGG TTTAAGCTTA ATTATCCGAG GTATTGCATA	12840
	TTTAACAACA AGACGAAACC TTGAAGAACT TGAGCCTAAT AATTATTTAG ACCATGTCAA	12900
	TGACGATGAA GAACTAATC ATACTGAGGC TGAAAAATCT TCTTCAAATA TTAAAGATGC	12960
15	TGAAAAACA GGTGTAGCTA CTGCATCAAC GGTGGTGGT GCTAAAAATG ATACTGAAAA	13020
	TACAGTGGCT GACGAACCAA GCATTCATGA AGGTACTGAA AAAACAGAAC CTCAACATCA	13080
	CATAGGTAAT CAACTGAAT CTAATCATGA TGAAGATCAT GACATCACTT CGGAGTCAGT	13140
20	AGAATCAGCm GaATCAGTTA AACAAGCACC ACmAAGTGAC gATTTaACAA ACGATTCAAA	13200
	TGAAGATGAA ATAGAGCAAT CATTAnAAGA ACCTGCGACT TATAAAGAAG ACAGACGTnC	13260
25	ATCAGTTGTA ATTGATGCAG AAAACATAT CGAAAAAGCT GAAGAnCAAT CTTCAGATAA	13320
	A	13321

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8549 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

40	ATGTGTTGTA AACTTTTATG TTGAAAAGC TACTTATCTC AATGAAAACA AGTAGCATTT	60
	AATAAATTAA TTAGTATACA GCTAGTTTTT CTAATTGTTC TTAACTTGA ATTAAGTTTG	120
	ACCGTATTAG AGAGGCAGAT TGATCCATCG TTTGAATTGC TTGTCCTTCA TTTTCGTTCA	180
45	AGCCATTACA AACAACTTCA AACTGTTGTG CCATTTGATC AAGACGCGCA TGAGCTTGTG	240
	TGTTTAAAAT AAACATATCG TCATAATGTG ATGGCGAATA GATAATTCGT CGTTGTATAC	300
	AAACGTATAA AAACCTTGTC ATATCAACGG TTTTGGCATT TTAAACCTC TGTGTTTTCC	360
50	ACGCATGTTT GCCCTTATTT AAATAATTTG CCCTTTTTTC GCCCCGAAAA AAAACACAA	420
	AAAAATAACC AACTTCCTAA ATTAATAGGT GGTGTGGTTT TGTTGATTGT AGGGGTATAA	480
55	AAATAACCGC ATTATTAAAG ATACGGTTAC TCTGTTATCT GTAAATATAA TAGTAGTTTA	540

	AAACAGGACT CCACATAAAA ATCAACTCCT TTATATACCA TAATGATACT ATATTTTCTA	660
	GTTTATTTCA ATTTTTCAGT TTTTAAAAAT GAGTTTCTGT TTTTATTTAT ACGCTTTTCT	720
5	GTTTTCTTTT TAAATTTTAT CTTTTTGTTA TTCCATTCAT TGTAAAATTC TATTAAATTA	780
	ACATAAAATT TTTCATGCCC TATTTTATTT GTTGATGAGA TATCAATGTA AAGACTCAAT	840
10	ATTGTTTTTA AATAGATTTG ATGCAACGAC TGATAAACCG TATTACTATC TGCTATGTTA	900
	TTGGTAAAAT GCATAGAAAA ATATTCTAAT TTATTCATGC AATATATATG GGTTTCATTA	960
	TACTTCTTAA TGAGTGATT TATACCTTGC AATACGTCAT TACTTTTAAT AACAATTTCT	1020
15	TTTTCACCTG TCGAAAAAGT CCACTGTTTA TCTCCTATAT TTTCTTTAAT TGTTTTCTTG	1080
	TTGTCAAATT CTAAAAATTAT AGCCCGTAAA CACTCTTCTT TATAATTCTC GTTCTTGAAA	1140
	GTACGAAGCA AAATTTTTAT AAATTCGGTA TTGGTGACTT TTTTATAAGT GTGATATTTT	1200
20	GCAATCTCTT TATCAGTAAA GACTGTTCTT AGTTCGTGAT TATCAAACT TAAATTCATC	1260
	TTATTCTCTA ATTCATTAAT TTTATCTTGC AAACCAACAT TTTCTAAAAT TTTCTTGTTT	1320
25	ATCTCCCCTA TATCAAACT CCTTTTCGAA ATTAATTTTG AAACTCGTC TGCCATTTCA	1380
	ACAGCCTTTT CTTTCCTTTT ATACCTTTTG TTAAATTTAT GAACCACCGT TGCAGCATAA	1440
	TACGATATCC CACCAGATAA AATAGATGAT ATTATCGGTA TGTATATATC ACCTTTTCATA	1500
30	TTTCCACCTC TTTTAACACA ATTAAGTATT ATGATACACA ACTTGCGCAA AAAGATGTAG	1560
	ACAGAACATA ATGGCGAACA AAAACAACCA CCCAGTAACT AGTATGGGTG GCGTAGACTA	1620
	TAACAACTCT ATGTTATCAA GATATATGTA TCGAGTGATG GCAAGGAAGA AGTCTCCTGC	1680
35	GGGACCAACA GTCAGATATA TGGCCTCTGC CGGGCTATAT AGTTCACTCC TACTATATAA	1740
	AAGTAAGTAT AACATAAAAA GCACCCCGTA AACTGTTATA CGGGAATGCT AAAGTCATAT	1800
	ATACTACGGG GAGTAGTATG AAAACTATGC TCTCTATCGT AAGAAAAAAC ACCCAGTGAC	1860
40	ATGCTTGGGT GAACAAGGAT AGATGTAAAT AGTTGATGCA TGTGTACACA TCATAACAAA	1920
	AAACTAGCCC GAAGcTAGCT ATAACATAAA AAAATAGGCA AGTACCGAAG TACCTGCCAG	1980
45	TTACGCACAT TTAAATCTTG AGAGTAATGT TAAAAAGTGT ATAGGAATAT TAACATCCAT	2040
	CCAAATAGTT ATTTAATAAC TGTAAGATTC CCTATAATTA ATGTAGCaAA ATTTTATTC	2100
	TAAGTAAATA CTAAATCGTG CTAAACTTAC CAAAACACT TATTCTATTA CCTGCCTTGT	2160
50	CTACCTCTCC TGTCGCTATA TAACGACGTT GTCCACTATT AGCAATATAA GTAATCCATC	2220
	TATAGCCATT GATGCAATAT GCGCCGTCAT ATTTAATTGT TGC GTTATTA GGTAATACAC	2280
55	CTGTAATTCT TGAATTAGTT GAATAGCCGT CCCTTACGTT ATTACCTTTA ACATTGGCAA	2340

	CTGGCACTGG	TGGATTTTTT	TGGTTTTTAG	CTGATGTTTT	AACATTACCA	GCTACCAAAC	2460
	CACCTATAGG	CTTACCATGA	ATCGCACCGG	CTATTAATTT	AGAATACAAG	TCATAGTTTT	2520
5	TCTTAATCCA	ATCCATATCA	TTTTTATTAG	TAATAAAACC	TAATTCAGAT	AAACGATAGT	2580
	TTATATTTAT	TTCTGCTGAT	ACATTAACTG	TTAGTAAATC	ATTACGAGGT	GTTACACCTC	2640
10	TTATTTGTCC	TAAGTTATTT	TTAATAACAT	CTTGTATACT	TTTATCAATA	GTATCTGCAT	2700
	TGAATTGACT	TGAAATAATA	ACATGCCCCA	CACTTGCACT	TTCTCCTGCT	GCGTCTAAAT	2760
	GAATCTCTAG	AACAATGTCA	TACCCATGTG	ATTTAACCCA	ATATAAGCCA	TAATCTTTAT	2820
15	TATTTCTTAC	ATTAACACCG	TAAGCAGTAT	CTTGATACAT	ATCTTGATGAT	TGACTTGAGC	2880
	CACCATATAA	TGCAACTTCG	TGACCTGCAT	GTCTTAAATA	CTTAGCGATA	TTTGGTGTTA	2940
	TATATTTACG	GATAAAATCA	CGTTCATTTG	TTCCGTTTCC	GACTGCTCCA	GGATCGTTAT	3000
20	AACCATGACC	GGCTACAAGC	ATAATTTTTT	TAGGTTTAAT	TACTGCTTGC	TTTTTGGCAG	3060
	TTGCTTGCTT	AATAACGCTT	TTAGCTTTAT	CTCCAACACT	TACTTTATCT	GGGAAATTTA	3120
25	ATCTAATAAA	ATACATTGGG	TCATCGTAAT	AATGAACATG	TCTTGTAACG	GTTTCGGGAC	3180
	CCCAACCAGG	TTGCGCAACG	CCATTTGTCC	AACCTTTACC	ATTCCAATTT	TGGCCAAACG	3240
	ATGTGAAAGT	GTTTAGATTA	GCGCTCTCAA	CAATTTCAAC	ATGTCCaGct	CCGCCACCAT	3300
30	ACTTTGACGG	GAAAACGACA	ATGTCCAAC	TTTGCGGTAA	AAAGCTATCA	TAGTTTTTTAA	3360
	TTATTTGCCC	GTATTTTTCA	ATCCTTGCTT	TATTATCAAA	TGGAATATTA	TAAGCGTATA	3420
	AACCTTGTA	CcTTTCGCCT	GTTGCTATCA	TAAAAACAT	ATTTGCGTAA	TCGTAACACT	3480
35	GAAATCCATA	AAACAAATCA	GGATTGAACT	GCTTCCCTAA	TGAATTATCA	AACCATTTTT	3540
	CTGCTTG GTT	TTTTGTTATC	AACATTGGTC	AACACCTACC	CTAAATCATT	TGTGTCGTTT	3600
	ATATTCGTAG	GTGTCAATTAC	TTCTTTAATT	GGCGCTTGCC	CTGTTGCTTT	TCTATACTTG	3660
40	TTTTCAGCTT	TATATTTCTT	TAGCTTTTGA	TTTGCCCAT	TACCTTCTTG	AGATGTTGGA	3720
	TTATCTTTAT	ATGTAGTATA	TAAAGCAACA	ACTGTTAAGA	TAATCGATGA	AACACTTTCT	3780
45	TCATCTACTG	GTATCGGACT	TATACCTTTA	TTGCTAAAA	ACTGATTGAC	TAATGCTAAG	3840
	ATCAATACGA	TGTATCTTGT	TATTACTTTT	GCATCCATTT	GTTTGCTCCT	TTTATCCAAA	3900
	ATAAAAAGCC	AGTGCCGAAG	CACTGACTCT	TAATATTAC	TTACACTTAC	TAAACCAGAA	3960
50	ACACGACCAA	AAGCTATATC	CTAAAATTCC	CTTAAGCATG	GTAATCACCT	CCTTTAAATG	4020
	CCAAAAATAG	TTTTTAACAA	GGCTATAACA	AATGTACTTA	GAATCGTCCC	TATTAATCCT	4080
55	AGAATCCACA	TCTTGATGTC	TCTAATATTT	TTAGCATTTT	TCTCTTTATT	TTTTTCATCT	4140

	TGCGTTCTCA	GACTGTCTTC	TATTCTGTCG	AATTTTTCAA	ACATAGTCTT	ATCATTCTCT	4260
	TCTAATCGCG	TTAAACGCCA	ATCTTGTTTCG	TGTCGTTTGG	TAAATCCAAA	CATTACACCA	4320
5	CCCACTTTAT	TCAAATTAAA	AAGCCATAAG	ATTATAACCT	ATGACTCTAG	ATTTTCTGGA	4380
	TACTTTTCTC	CTGTAATAAT	TGCATATTCC	TCTTTATCTA	TAACTTCCAT	ATCTACATAC	4440
10	CACGCTATAT	CTTCTTTACT	ATATTCTTTC	AATTGATACC	ATGTTTTAAT	ATCTTCGAAT	4500
	GTTGGTGAAA	TTAATTTAAG	CATTTTCAGT	CTCTCCTTTA	ACCTCTTCTA	ATTTTTTATT	4560
	AAGTGTCAAC	AGTTGTTTTG	CCATTAGTGC	ATTTTGCTTA	TTAACTTGCA	TCGATAACTT	4620
15	TGTACTTTGA	ACAAC TTGTT	TCTGCATACT	AGCAACCATT	TTTCGTAAGA	TGTCATCAGA	4680
	AGCGACTGTG	TTTTGTTCTT	CACTGTCAAT	CTGTTGATGC	AAGTCATCTT	TTTCTTCTGA	4740
	ATAATCTTCG	TTAAAACTA	TTTCCCCATT	TGAATATTTA	AAGGCTTTAG	GTCTAAAAAC	4800
20	TTGAGAGAAA	TTTTCTGGTA	AATTTTCAAT	ATCAATACCT	TCTTCAAAGC	CACCAATGAT	4860
	AGCGTATGAA	ATTATCTCAT	TACGCTTGTT	AAC TAATATT	TGCATTATTT	TCTCACTCCT	4920
25	ATAATTTTGT	TAATTGTCCC	TCTATTTGCG	TTCGCACCAG	AGCCTCTTTG	ACTTCCTAAG	4980
	TCGAAATAGA	CATCGTTTGA	TATAGTTAAA	GATGTACGAC	TAGATTTAGT	TAATCCAAAC	5040
	TCATAAACAC	CTCCACCATT	TCCATCACCA	TCTGGAAGAT	TTGAGGGATT	CAATGAAATC	5100
30	TTTCCTCCTC	CAAAGGACT	GCCAACTCT	GTAAAGTCAC	CACCTGGAAA	AGTCCCATAA	5160
	AAAATTAATA	AAATAAATTG	GTCTAAACTC	TCATTTAAGT	ACAATGTAGA	GCCCACACCA	5220
	TTTGCTGTTT	CATCAAAAAT	AACCGAATAC	CTTTTATTAA	ACTTGTCATC	TGCGTATAAT	5280
35	TTAGCGTTAC	TTTCGGCCAT	ATTAGCTTTT	GATTGGGCAC	TTTGAACAGT	TTCAAAGGT	5340
	GTATTGTAAT	CATTAATAGC	TAATTCTGAC	CACTCAGACC	ATGAACCCGC	TTCTTTTCTT	5400
	TTAAACAAATA	CTTTATTTGT	ACCGTTCGGT	CGATAAGTCA	TACGCTTGTA	ATCTGAAGTT	5460
40	ACTACTAAAT	ATTCGACAGT	ACCGTTAGTA	CTAACACCTC	TTGGATAATT	TATAGCTTGC	5520
	GAAACATAAA	TAAATTGGGT	TGAATCACCT	ATTCTTTGTT	CTGGATTATT	AAAATCAAAT	5580
45	CCAGTAATCT	GCATTATCTT	ACCATCATCT	TTAGTAATCT	TAGCTTTTTG	CCAATTTGAA	5640
	GTAGAACAC	TTGTGACTAA	ACCACCACTA	TTCACTGACT	GCTTGAAGGC	TTCATGTTTC	5700
	TCATCCATAT	ATCGCTTTTG	CTCATCGAAT	GTTCTTGAAT	ATGCTTGCGC	TTTATTTTCC	5760
50	AAATCAGATA	TATGGCTATT	AGCAAGTTGC	TTTAATTCAT	CTATACTTGA	AGATTTTGCT	5820
	ATTTGAATAT	CTGATAGACC	TTTTTCTTTA	GCTTTTTTCAA	TCAGACTCGC	ATAATCTTCA	5880
55	CCATTTTTTA	TAGCCTCGTC	CATTGCTTTC	GCACGATCCA	TAATAGTTTT	TTCTAATTCC	5940

	TCAACGTTAA	ATGTGATAGT	TCTCTCGACA	ACTACCACGT	CTGAATTACC	TAATTCTGCA	6060
	ACCGAAACTT	GAGCTTGATA	ACTTCCATCT	CGTTTAATTA	CATCATTAGG	TAATTGAAAT	6120
5	TTTAAAATAC	CTTTAAATGG	ATCTAATATT	TCTAGTGGAG	CAACTACCAT	GACTCCTTTA	6180
	CCTCGAATCG	CTATTCGTGC	KTTGATATTT	tCTTCACTCA	ATAATAACGG	TTGATTATTT	6240
10	TTAGTGATAT	TAAAAAGAAG	AACAGAAGAA	TCACTCTCTC	CTGTTCTAAA	AGTTATATCT	6300
	AGATTTGAAA	TATTTCCATA	ATGCGCTGTG	TTTTCTAAAT	TTATAGCTAC	AGATTTCTCT	6360
	AAATTACTCA	TTAACTTATA	ATTCTCCCTT	CGTGTAAGT	CCATGGCCCT	GAACCTGTTT	6420
15	TACTATCATA	ATTTTTCAAT	AGTATCTCAG	CAGATGCTGT	AACACTATTA	CGAACTAGCC	6480
	TATGAACAAA	GCCACCTGTG	TTTGAAGCTT	CTACATATAA	GTTCCAACCA	GCTACCCCTT	6540
	TACGTTCACT	TGGAAAATCT	GTAAAACGTT	TTGTATCATC	CGTAGTTAAA	TAAAACGACA	6600
20	TGCCTACTAT	GTTAATATCT	GACATTTTTG	TGATGAATGA	AGGTACTCTC	TCCCATTTAC	6660
	CACTATTTTT	AGGCACATAA	TTCCAGTCCG	AAATGTCTCC	AGTTCTTCCA	GAAAGCACCC	6720
25	TTTCAAAGT	CATCATATTC	CTTGCATAAC	TATTACGCGT	CAATATCTGA	ATTACATCAC	6780
	CGCCAGTTTG	TGGTGGCTTA	ACTTCCAAGA	ACCAACCTGC	ATCACGCCAT	TCTCTTGGTA	6840
	ATGGGAAATC	ATCGATTTGA	ACTGTATGAT	CAGTGTATAA	ATAGTAAAGA	CCTGGCTCTG	6900
30	TTAACATCCC	AAGATTCTTA	AGTTTATCAG	GCCTCATTGG	TAAAGGTTTA	ACTCTACCAC	6960
	CTGTGTCACT	CaTGATAAAA	GGAACGCCTC	TTGAGTGAAG	TATTTCTAAA	ATACCTCTTT	7020
	GCCCAATCAT	GAAAATACGA	TGTGTTCTAT	TTCCaTCACC	ACCGACAGTA	ACACCTAGCA	7080
35	TCAAAGCTTT	TTTACCACTA	TCTTTGTCAT	AGTATATTTG	CAAACCTTtC	TgCTTCCGCA	7140
	AATTCGCCAG	GAAATGAATC	tAgTGTTCCA	CCATAGTCAG	CATTAACCTG	ATACGCTTCT	7200
	TCTCtGTTTT	CTAAATCGAA	AGCCGTTAAA	TAGTTTCTAT	TATTTGGATT	ACTGTCTCCT	7260
40	GTATACCAAT	ACAAGTATTT	TTCATCAAAA	GTCACACCCT	GCATTGGTTG	GGTTTCGTTT	7320
	GTTAGTCTCA	TAGGGATACT	GATTTTATGC	AAAACTTTAT	CAATATTTTT	ATCAACATCG	7380
45	TCTAAACTTC	TTATCTCTAT	ATAAnTCATT	GAGTTTTCAA	GTTCCCACTG	ACTTCTAGGT	7440
	CTCTCaATTC	TGTATAGAAT	TTTATTTTCT	TTTTCATTTA	TGACAGGGGT	GATGTAGGGT	7500
	TTTTCTGGGT	GTCCTGTAAA	TACATCTTGC	ATACCATACT	TGCCATAGCT	AATTTCCACA	7560
50	TTAGGCGTAT	ACTTGAAACG	AACTAATGTA	TTCTCATTAT	TACCATTTAA	GATAAACTA	7620
	TAAATCCATA	ACTCATcATC	AATATATCTA	TAACCGTTAT	GTGTACCATG	ACCCCCACCT	7680
55	ACAATCAATG	AGCTGTCTAT	AAATTGACCA	TTAGGTCTTA	GACGACTTAG	CATATAGCCA	7740

ATTACTGCAT TTGTAAgAGG TGCAAGTTCT GTCACAAATA AAAATTCTTG CTTATCAGGT 7860
 TCAAAACGAT ACTCGATATC AAGAATTTCT TGTTTGGTCT TATTTAATTC TCTTATAGTT 7920
 5 TCCTCTTTAT TAATTTGAGT TTTGGTTTCC CAATCGTCTA AATGTTCTTT TAATGTGTCA 7980
 AAGGTTTCGC CGTTTACATT AACTCGAGCT TGAACAATCT CATTAGCACT GTTATTACGT 8040
 10 GGTGCCACAA CAAGTGC GTT AATTTGACTT TGTAAGATT TGTTTACTGC TGCTTGCGAT 8100
 CTACCATTAT AATAAATTTG CTCAGCGAAG TGTTGAATTG TTTTAGCTyT CTGATGCAAC 8160
 TTAAACTCTG TTGTCAAGCC AAGCGCAAAT TGCTCTATTCT TTTGTAAGTT TTGTATTTCC 8220
 15 TTAGCTCTAT AATCTCGACC TGCTAAAGCT CCCAAATCCT TTATTAAATA CAAATTTTCC 8280
 ATAATGCACC TTCCTTTCTA ATAAAATAGC ACTGTACCAA GTTCCCACT ATCGTCAACT 8340
 GTTATTTTCC ACAATTTACC GTTTGGGGAT TTCTGTACAA TGCTATTTTG AATAATTgcC 8400
 20 TGctTCGCCT ATTTTAAAT TATCTAATTT ATTTkTATCA TTTACCGAAA TGATACCGTC 8460
 TTGAGGCAAT CCATCAATAn CACTACTGCC TGCATAAGGT ATCCCATTTA TAGCTTTCCA 8520
 25 ATGTGTAGCT GGAAAGTACT GTTTATCGT 8549

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 3601 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGGCGTGTAG TGA CTTACGG nTAGGAACT ATGTATCCGA ATGATTTATT GAGACCAAAA 60
 AGGCATTAAA GTCCATTGAA ATATCnGGTA GCGmGTTGGT ACgTGGACGT GGGGGCCCTA 120
 40 GATGTATGAG TCAACCATTA TTCAGAGAGG ACATTTAACG TAATAAATTA TAGAmACGAG 180
 GGTGAAAATA ATGACAGAAA TTCAAAAACC GTATGATTTA AAAGGCAGAT CATTATTAAA 240
 45 AGAAAGTGAT TTTACCAAAG CAGAATTCGA AGGACTTATT GATTTTGCAA TTACATTAAA 300
 AGAGTATAAG AAAAACGGTA TTAAGCATCA CTACTTATCT GGAAAAAATA TTGCACTACT 360
 ATTCGAAAAG AATTCGACGA GAACGCGTGC TGCGTTTACA GTTGCGTCTA TTGATTTAGG 420
 50 TGCGCATCCA GAATTTT TAG GAAAAATGA TATTCAATTA GGCAAAAAG AATCTGTAGA 480
 GGATACTGCG AAAGTATTAG GTAGAATGTT CGATGGTATT GAATCCGTG GTTTTTCACA 540
 55 ACAAGCTGTT GAAGATTTAG CGAAGTTCTC TGGTGTACCG GTGTGGAATG GATTAACAGA 600

	TCTAGAAGGA ATAAACTTAA CTTACGTTGG AGATGGACGT AATAATATTG CGCATTTCATT	720
	AATGGTAGCA GGTGCTATGT TAGGTGTTAA TGTAAGAATT TGTACACCTA AATCATTAAA	780
5	TCCAAAAGAG GCATATGTTG ATATTGcAAA rGAAAAaGCG AGTCaATATG GTGGTyCAGT	840
	CATGATTACG GATAATATTG CAGArcCAGT TGAAAaTwCm GATGCTATAT ATmCAGATGT	900
10	TTGGGTATCG ATGGGTGAAG AAAGTGAATT TGAACAcGTA TTAATTTATT AAAAGACTAT	960
	CAAGTGAATC AACAGATGTT TGATTTAACA GGTAAAGATT CAACGATATT CTTACATTGT	1020
	TTACCAGCAT TCCATGATAC AAATACACTT TATGGACAAG AAATTTATGA AAAATATGGA	1080
15	TTAGCTGAAA TGGAAGTTAC AGACCAAATC TTTAGAAGTG AACATTCAAA AGTGTTTGAT	1140
	CAAGCTGAAA ATAGAATGCA TACAATTAAG GCAGTAATGG CAGCAACATT GGGGAGTTAA	1200
	TCACTAAATG GAACGATATG AATATGATGT GTCTGATGAT ATAAGTGTCa TGTACAGACA	1260
20	CCTCATATTG GTATTAAAGG AGAAATGAAT ATGAACGAAT CAGGAGATAA CAAACTCAGT	1320
	AAATCTTCTT TAATTGGACT AGTTATAGGA TCCATGATTG GTGGCGGTGC GTTCAATATA	1380
	ATGTCTGATA TGGGCGGTAA AGCCGGTGGA TTAGCCATTA TTATTGGTTG GATTATTACA	1440
25	GCTATAGGAA TGATTTTCATT AGCGTTCGTA TTTCAAaATT TAACCAATGA ACGGCCGGAG	1500
	CTAGACGGTG GTATTTATAG TTATGmTCAA GCAGGATTTG GCGATTTTGT AGGATTTATC	1560
30	AGTGmTTGGG GATATTGGTT CTCAGCGTTT TTAGGCAATG TTGCCTATGC AACACTATTG	1620
	ATGTCAGCAG TAGGTAACCT TTTCCCGATT TTTAAAGGAG GCAACACATT ACCAAGTGTT	1680
	ATTGTCGCCT CGTTACTACT CTGGGGTGTC CATTTCTTGA TTTTAAAAGG CGTTGAAACA	1740
35	GCAGCATTTA TCAATAGTAT TGTTACTGTT GCAAAGTTAA TACCGATTTT ACTTGTAATC	1800
	ATATGCATGA TAATTGCATT CAATTTTGAC ACTTTTAAAA CAGGCTTTTT CAGTATGACG	1860
	TCAGAGGGTG TATTGCCATT TAGTTGGGCG AGCACAATGA GCCaAGTtAA AAGTACGrTG	1920
40	CTAGTGACAG TTTGGGTGTT TATCGGTATC GAAGGTGCAG TAATTTTTC TAGTAGAGCT	1980
	nAAAATGAGA AAGATGTAGG TAGTGCCACG GTTATAGGAC TTATATCAGT TTTAATTATC	2040
45	TATyTCTTAT TAACTGTATT AGCTCAAGGC GTGATTTTGC AAAATCATAT TTCGCAATTA	2100
	GATTCGCCAA GTATGCCACA GGTGCTTGCA ACTATTGTAG GTGGTTGGGG ATCTACACTT	2160
	GTAAATATTG GTTTAATTAT TTCGGTACTA GGTGCATGGT TAGGATGGAC ACTGCTTGCT	2220
50	GGTGAATTAC CTTTCATTGT TGCAAAAGAT GGATTATTTT CAAAATGGTT TGCTAAAGAA	2280
	AATAAAATG GAGCACCTGT AAATGCACTG CTTATTACCA ATATATTAGT ACAATTATTT	2340
55	TTAATAAGTA TGCTATTTAC ACAGAGTGCG TATCAATTTG CATTTTCACT AGCATCAAGT	2400

CGACAGCAAG CAACTACTAA ACAATGGACG ATTGGTATCA TAGCCTCAAT TTATGCTATA 2520
 TGGCTTATAT ATGCAGCAGG TATCAATTAC TTATTATTGA CGATGTTACT TTATATTCCA 2580
 5 GCTCTTCTTG TTTATACaAT CGkTCmAAAG rATwATCAGa CACGTTTGAT TAAATCAGrC 2640
 TATATTCTTT TTATGATTAT tATCGTACTT GCAGTTATCG GGTTAATTAA GTTATTGATG 2700
 10 GGAACGATAA ATGTTTTTTA AAAGGAGCGA CAAAAATATG AAAGAGAAAA TTGTCATTGC 2760
 ATTAGGCGGT AATGCGATAC AGACAACAGA AGCAACAGCT GAAGCACAAC AAACAGCTAT 2820
 TAGATGTGCG ATGCAAAACC TTAAACCTTT ATTTGATTCA CCAGCGCGTA TTGTCATTTC 2880
 15 ACATGGTAAT GGTCCACAAA TTGGAAGTTT ATTAATCCAA CAAGCTAAAT CGAACAGTGA 2940
 CACAACGCCG GCAATGCCAT TGGATACTTG TGGTGCAATG TCACAGGGTA TGATAGGCTA 3000
 TTGGTTGGAA ACTGAAATCA ATCGCATTTT AACTGAAATG AATAGTGATA GAACTGTAGG 3060
 20 CACAATCGTT ACACGTGTGG AAGTAGATAA AGATGATCCA CGATTTGATa ACCCAACTAA 3120
 AcCAaTTGGT CCTTTTTATA CGAAAGAAGA AGTTGAAGAA TTACAAAAAG AACAGCCAGA 3180
 25 CTCAGTCTTT aAAGAAGATG CAGGACGTGG TTATAGAAAA GTAGTTGcGT CACCACTACC 3240
 TCaATCTATA CTAGAACACC AGTTAATTCG AACTTTAGCA GACGGTAAAA ATATTGTCAT 3300
 TGCATGCGGT GGTGGCGGTA TTCCAGTTAT AAAAAAAGAA AATACCTATG AAGGTGTTGA 3360
 30 AGCGGTTATA GATAAAGATT TTGCTAGTGA GAAATTAGCA ACGCTGATTG AAGCAGATAC 3420
 CTTAATGATT CTTACGAATG TAGAAAATGT ATTTATTAAC TTTAATGAAC CTAATCAACA 3480
 ACAAATCGAT GATATTGATG TAGCAACACT GAAAAAAtAC GCGGCACAAG GTAAGTTTGT 3540
 35 GGAAGGATCG tGTTGCCAAA AATAGAAGCT GCGtACgtTT GTTGAAAGtG GGGaAACCAA 3600
 A 3601

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGACACTATT AAATGAATTA GAGCACAATC TAACAAATCA AATTCATTTT TCAAAAGATG 60
 AACGACTCAC ACATATCGCT TTAAAGTTAT TCGAAACAAC CGATCCTGTT TCAACAAAGC 120
 55 AACTTGCGCA AGATGTTAAT GTTTCGCGTC GGACAATTGC AGATGATATT AAAATGATTC 180

TTATTGGTGA GGAAGATCAT TATCGTAAAG CGTATGCACA CTTTATACAT CAATATATGA 300
 AACAAAGCTGC ACCTTTTATA GAGGCGGATA TCTTTAATTC AGAATCAATC GCATTGGTTC 360
 5 GCCGTGCCAT TATTAAGACA TTAAATAGTG AAAATTATCA TTTAGTTCAG TCGGCTATCG 420
 ATGGCTTAAT CTATCATATA CTCATTGCCA TTCAGCGTTT AAATGAAAAT TTTTCGTTTCG 480
 ATATACCTAT CAATGAAATT GATAAATGGC GACATACTAA TCAGTATGCh ATTGCTTCAA 540
 10 AAATGATAGA AAACCTAGAA CGCAGTGTA TGT 573

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1221 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTGATATTTA TAACGTTATA TTTTAATAGT TCACCTGGAT TATTAAATAA ATAGTCCGCC 60
 25 AAATTTTCTT TTTCTTTATC AATCTGaTkG TAATTAACaC TTTCGaCTTC TGTAGGAATT 120
 CTAATGTCAA CAGAAGCATT GATATAAGCT TGATGTTGCA TGCAATCACA CTCCTAATCC 180
 30 TTCATmTmAA ACGGAGAAGT AAACCCGTCA CTATTCAAAT TCAATCCTTT TGCCCAATCA 240
 ACAGGCTTAT TCATGATAGT TTCGATTTCC TTAAGTCCAT TTGAACCTCT AGGTATTTCT 300
 ACAATTACTT CATCATGGAC ATGGCCAACT ATTTTAAAAC CTAATGCTTC AAGCCTTGCT 360
 35 ATAGAAATCG CAAGTAAATC CCTTGCAGTT GCTTGAACAA TATTCTCGAC TAACTTCCCA 420
 CCATACGTTT TTAACTTTGA CCATTTACGG TTAAGATCTA ACCCCATAAA TTCAACAACT 480
 TGACTACCCC AACTATTTTC ACCAACTAAA GCTTTTGGAT AAGCTAAAGC TCTTCCACTA 540
 40 GGCAGTTCAA TCATTAGAAA ACCTTTTTTC ATATAAAATC TAAGTCCATG TGTATGATGC 600
 GTCTTTCTGGG ATTTTACAGT ATTAATTGCA GCCTCTTGGC AAGCCTTCCA AAAATTAACT 660
 ATGTTAGGAT TTGCGTTACG CCAACTATCA ACTAAACCTT GTAACCTGTT TTCTTCAATG 720
 45 CCCATTTCCA ATGCACCCAT TGCTTTTAAA GCTCCAGCGC CACCTTGATA GCCTAAAGCT 780
 AATTCGGACA CTTTTCTTTT TTGTCTGAGA GGGTCGCCTT TAGTTATGCT TTCTACCGGT 840
 50 ACATTAAACA TTTGAGAAGC CGATGCTTCA TATATCTTTC CGTGTGTGTT GAATACATCT 900
 AAACGCCATT GTTCTTTTGC ATACCATGCT ATGACTCTTG CCTCTATTGC AGAAAAATCA 960
 CTTACTGCTA GTTCATTACC TTCTTCAGCA GTAAATGTCG TCCTAACTAA TTGACTTAAT 1020

AGATCTCTTG CTATTTCTAA TTCAGTATCT GAAATATAAT GCTTTGTAA ATTCTGAAGT 1140
 TGTACACCTC TACCTGCCCCA TCTTCCAGTA CCGGCACCGT AAAATTGAAA CAGACCTCTT 1200
 5 ACCCGTTCAT CACTGCACAT C 1221

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1090 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTTGTTTGG TATGAGGTAG CAATGACGAC GTGTCATTGG TGGAGATTGT AAAAATACAT 60
 20 AATAAAAAGA AGCGGCAATG TATACCGCTC CTTTTTTATA CTACATACCG ATTTTCAACC 120
 ATCTCTTTCT ACTTAGTAAT AAGACAATAG TATTAACTAT AAATAGAAGA ACGAAGAATG 180
 ATACTATATT TATAATTTCA GTAGGACACA TAAATGTTGA CTCGTTATT C AATATTTTTT 240
 25 CTACGGCACG ATACATCGTA TTGCTCGCCT CAAATGGAGC AACGATACCA AATATATTTT 300
 TATTAATGGC AACTAAGATG ACTGAACCAA TCCAATATAC AATGCTGATA CCTAAGCTGA 360
 30 TTAAATGTT AGGTGAAACC ATACTAATCG TTCCAACAAC TAAGATATAT TGTAAGATAA 420
 CGAGTGAAAA TAAGATTATT AATAGTAAGT AATGTGAGAA ATCCGAATAT ATAATTGAAA 480
 TAATAGTGAT ACTTAGAATT ATGAACACTA AACATTCAAA AAATAACACT GCTACCTTTT 540
 35 TATAGAAGAA GGTAAAGATA TTATCGCCAA TCAATTTATA AAACAGGATA TTTTATTTCG 600
 AATACTCTTT ATTAATAAAA TATGCAATAA CAAATGAAAA TAGTAAGAAC CCTAATTGCG 660
 TTGCAACAGT ATATGAACTG AAGAAAAACT GGCTATAGCT TAACTTTTA ACTTTGTCTA 720
 40 TACCTATTGG TAAAAAATAC CCAAGTAAGA AAAGGAATGT GAATAGCACA ACAAGCGTGT 780
 AAATAATTTT ATTGGAAATA CTTTTTTTAA ATTCTAATTT CAAAGTGGAC ACCTCAATTA 840
 TAAATTAATG TAATCATTTA TGACTTCTTC TTTTGATTGG TACTCTTCTA TTTGAAGGTC 900
 TTTAAAAATA AAGTATTTAC CCGGCAAAGC ACTTAAATCG GATAAATTaT GTGTAATATT 960
 GATAATAGTT TTAGTTTGAT GGCTTTGAAT AAAATCATTT AAAAATTCAT AAATTTCAAT 1020
 50 AACTGTTTTT TGTCTAAAG CGTTTGTAAC TTCATCTAAT ATGATTAAAT CATGATCTTC 1080
 CAATAAGAAA 1090

(2) INFORMATION FOR SEQ ID NO: 10:

(A) LENGTH: 904 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```

10  TTAGGACTAT TTTATCATAT TCATTTTAAAT TACGGCTAAA AATTTTAAAA ACGGGGATTA      60
    ATATATGGAA TTAAGCTATG AAAGTTAATT GATACTTGCA TTTTACGCTG ATTTATATAA      120
    GAATAACTAT TGTATAGTTT TAAAAACGAA CGTACGTTTG CAGGAGGCGA AATCATTGGC      180
15  AATGAATAAA CAAAATAAATT ATTCAGATGA TTCAATACAG GTTTTAGAGG GGTTAGAAGC      240
    AGTTCGTAAA AGACCTGGTA TGTATATTGG ATCAACTGAT AAACGGGGAT TACATCATCT      300
    AGTATATGAA ATTGTCGATA ACTCCGTCGA TGAAGTATTG AATGGTTACG GTAACGAAAT      360
20  AGATGTAACA ATTAATAAAG ATGGTAGTAT TTCTATAGAA GATAATGGAC GTGGTATGCC      420
    AACAGGTATA CATAAATCAG GTAAACCGAC AGTCGAAGTT ATCTTTACTG TTTTACATGC      480
    AGGAGGTAAA TTTGGACAAG GCGGCTATAA AACTTCAGGT GGTCTTCACG GTGTTGGTGC      540
25  TTCAGTTGTA AATGCATTGA GTGAATGGCT TGAAGTTGAA ATCCATCGAG ATGGTAATAT      600
    ATATCATCAA AGTTTTAAAA ACGGTGGTTC GCCATCTTCT GGTTTAGTGA AAAAAGGTAA      660
30  AACTAAGAAA ACAGGTACCA AAGTAACATT TAAACCTGAT GACACAATTT TTAAAGCATC      720
    TACATCATTT AATTTTGATG TTTTAAGTGA ACGACTACAA GAGTCTGCGT TCTTATTGAA      780
    AAATTTAAAA ATAACGCTTA ATGATTTACG CnwGGgTAAA GAGCGTCAAG AGCATTACCA      840
35  TTATGAAGAA GGGATCaAAG rGTTgTTAGT atGTCCAaTG ArGGAAAAGA AGTTTTGCCT      900
    GACG                                          904
  
```

(2) INFORMATION FOR SEQ ID NO: 11:

40

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11271 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```

50  GATTTCTAAA TCAAGATCTG TTTTACGATA ACCATTCAAA CCTTGACGTT CATCTTCTTC      60
    AGGTTGATTT TGTTGCTGTG TGTCTTTGTT GTCAGAAGTC GCTACTGTTT TTTTATTATC      120
55  TGTTTCTTTA GTCATAACAA ACGCCTCCGT TATAAAACGC TATATTTAAT GATATGTGAT      180
  
```

	TTAATAAGAC GATTCAGCAA GTTTTAAAGT ATTATTTGAC TATGTTGGAT TAGGCATCTA	300
	GTCCCTATAAT ATCACTGACA TTGTCAAAAT GATGATCTTT TAAGTAACGT GCGATGCCTT	360
5	TGTTTCATTTT CTTAGTTAAA CCTGGGCCTT CAATAACAAG TGATGAATAA ATTTGAATAA	420
	GTGACGCACC GTGACGCATC ATTTTGATTG CATCTTCAGT ACTGAATACG CCGCCTGTAC	480
10	CTATAATTAA AAATTCACCA TTTGTTTGCT GATAAgCATA CTTAATCAAT TTTAAATTAC	540
	GTTCAAATAA TGGACGACCA CTCAAACCGC CTTCTTCGAC TTTATTAGCA GAAGTTAAAC	600
	CATCTCGTTG TCGCGTTGTG TTTGCTAAGA TGATACCGTC AAATGTCTCA GTAATCGCTG	660
15	GTAATAGTGC TTTTAAGCCA TCGAAATCCA TATCAGACGT TAGTTTTAAA TAAATTGGCA	720
	CTGTTACATC ATGTTGTTTT TTAAATGCTG TTAAAGCTTG GCATAACATT GAAAATTCAT	780
	CTTTATCATG GAAGTTTTGA AGATTTTCAG TATTTGGAGA ACTGATGTTG ACTGTGAAAA	840
20	ATGAAACGTC GTGTTTAAAC GTATCAATAA CCTTTATATA ATCTTGATAA CGCGCTTCAT	900
	AAGGTGTCAT TTTATTCACA CCAACATTGA TACCAACAGG TACTTGATAA GCATTTTTAC	960
	GCAAATGACT TAGTGCTTTG TTCATACCAA TATTATTGAA GCCCATTCGA TTTATCAAGG	1020
25	CGTCATCTTC TAATAATCTA AACATGCGTG GTTGAGGGTT ACCCGGTTGA GGTTTAGGTG	1080
	TGATACCACC TAATTCTAAA GCACCGAATC CAAGGTGTTT CAATGCTTTT GGTACTTCGC	1140
30	AAGATTTGTC GAAACCAGCT GCTAAgCCAA TTGGATTGTC GTACGTATTA CCTTGTATCG	1200
	TTTGTGATAA CGTTGGATTG TTATAAGTAA ATAGTTTATC GACGACTGGG AATAAAACCG	1260
	GaAACTTTTG TaACGTTTTT AATGCATCGA TAGTTAGTCC GTGTGCTTTT TCGGGTTCGA	1320
35	TTTTGAATAA GAAAGGTTTA ATTAATTTGT ACATGAGTAT GTCCTATTT CATTATATTT	1380
	GAGGCTTACT ATCCTCAACT TAATATATGT GAAATATATT CTTTAAATAG ACTAGCATT	1440
	CCATACATAA TTTCTAGTT AAAACTAAAA AGTTTTGAAA ATTGACGCAA gTTTGAATAA	1500
40	CGTTTTTAAG ATTAAATCAT CCTAATTAGG CAATATTATA GTATAAAGTA AGTAGATTGG	1560
	AAGGTGTTTG TATGAATGAA CAATGGTTAG AGCATTTACC TTTAAAAGAT ATTAAAGAGA	1620
45	TTTACCAGT GAGTGGTGGT GATGTAAACG AAGCATATCG AGTCGAAACA GATACGGATA	1680
	CATTTTTCTT ACTGTCCAA CGTGGACGTA AAGAATCATT TTATGCTGCA GAAATTGCAG	1740
	GTTTAAATGA ATTTGAACGT GCAGGTATCA CGGCACCTAG AGTAATTGCA AGTGGCGAGG	1800
50	TTAACGGTGA TGCCTATTTA GTGATGACGT ATTTAGAAGA AGGGGCTTCA GGGAGTCAAC	1860
	GCCAATTAGG GCAACTCGTA GCTCAATTAC ACAGTCAGCA ACAAGAAGAA GGCAAATTTG	1920
	GCTTCTCATT ACCTTATGAA GGTGGCGATA TTTCTTTTGA TAATCATTTG CAAGACGATT	1980
55		

	GGCTATGGGA	TGCCAACGAT	ATCAAAGTAT	ATGACAAAGT	GCGACGTCAA	ATTGTGGCGG	2100
	AATTAGAAAA	GCATCAAAGT	AAACCGTCTT	TATTACATGG	TGACCTATGG	GGTGGTAATT	2160
5	ATATGTTCTT	ACAAGATGGT	CGTCCGGCGT	TATTTGATCC	AGCGCCATTA	TATGGTGACA	2220
	GAGAATTCGA	TATCGGTATT	ACAACGGTAT	TTGGTGGTTT	TACGAGCGAA	TTTTATGATG	2280
10	CGTATAATAA	ACATTATCCA	CTCGCAAAAG	GTGCATCCTA	TAGACTTGAA	TTTTATCGTT	2340
	TATATTTATT	GATGGTCCAT	TTATTGAAAT	TTGGTGAGAT	GTACCGTGAT	AGTGTTCGCG	2400
	ATTCTATGGA	TAAGATTTTA	CAAGATACAA	CAAGTTAGTT	AAGACGTTAG	ATTGAGATAA	2460
15	ATAGATAATA	TGCACAGATA	TTTTTACAAT	GAGAAGCGAT	ACAGCTGCCT	CAATAAAAAAT	2520
	ATTTGTGCGT	TTTTATTGTT	GGAAAATAAA	ATTTTAATCG	CTATTGTTAA	TTTCTGTAAT	2580
	GTAAAAAAG	GTTGAGTTAC	AATAAAAGTG	ATTTTATAAC	TTTTTGTTCA	ATAAAATTCT	2640
20	AGGAATGATA	CATATTTATT	GATACAATAA	TTTTGAATAT	AATCATAAAA	CAATATTTAA	2700
	GTATAATTGA	ATGTTTGAAT	ATCATATATT	GATACAGTTT	CTAATAATTT	TAAAAATAATT	2760
25	TAAATGGAGA	GAGGTGTAAA	TGATGAGTAC	AGTTCAAAGT	GATATTTTTA	AGACCAATAG	2820
	TGCATCATCA	TCTATTAAAA	GCGCTGTTGA	AACATGTAAT	AATGTGTCGA	AACCGGATAA	2880
	AGATGAAAGT	ACAACAGTAA	GTGGAAATAA	TAATGCTCAT	AGTGTGATAG	ATGATTTGAT	2940
30	GAGTAAGAAT	CAATCTGTTG	CTGAAGCAAT	ACGAACTGCG	AGCGATAATA	TACAAAAAGT	3000
	TGGTGAGGCT	TTTGACCAAA	CTGACGTAAT	GATTGGTAAT	GAAATTGGTA	AAAATTAAAA	3060
	CGTGGTGAAA	TGATGTCGAA	TAACTGGAT	GAAATCAATA	AAATAATCAC	AGCGAAACAT	3120
35	GAGCAAATGG	ATGACTTATA	TGATGAAAAG	CGAGAGGTTA	AAGCATTGAT	AGATGAAAGT	3180
	GATGCGCTTA	ATCATTCGAT	AGATCAATTA	TATCAACATT	TAGGTGAGCG	TTATTATAGT	3240
	AGCAATATGG	CTAGTCGTAT	GGAACAGTTC	CGCGATGAAT	TTCATTTTGC	GAAACGACGT	3300
40	TCAACGGAAG	CGTTATACGA	GCAGCAACAG	CAAATTCAAC	ATGGCATTCT	TAAAGTGGA	3360
	GAAGAGATGA	TTGACTTGGA	AATGCGAAGG	AATGTTGAAA	TTGAGACGGT	GACAAAGGAG	3420
45	GAAAATAAAT	GGAAACAATA	GGAAGCATT	TTTATTTAAA	AGAAGGTTCT	CAAAAGTTAA	3480
	TGATTATTAA	TAGAGGmCCA	aTTGTAGAAA	TTGAAAATCA	AAAGTATATG	TTTGAATATT	3540
	CTGCATGTAA	ATATCCGATT	GGTGTGTAG	AAGATGAAAT	TTATTATTTT	AACGAGGAAA	3600
50	ATATAGATT	AGTTATTTTT	AAAGGTTATT	CTGATCAAGA	TGAGGTTAGA	TTTCAAGAGT	3660
	TGTTTGAAAA	TATGAAACAA	AATTTGGATA	GTGAAATACA	ACGTGGAGAA	GTTACACAAC	3720
55	AATAAAGAAA	TACTTTTTCT	TTATTGGGGT	GGGACGACGA	AATAAATTTT	GTAAAAATAT	3780

	ATGTCATTCA TAATCATTTG AACTAAACGT AGCAGCCTTA AATTTTAAAA AAAGACACAT	3900
	ACCAACTTCC GAAATGTAGA TGAATTCTCT ACAATAACGG AAGTTTTTCT TTTAATATTG	3960
5	AAATTTCTCA AGGATAGGTC TATACTTTAT AAATCGTAAT TATTACGATT TATAATCAAA	4020
	AACAATAACT TGAAATAGAT CATTGAGGGA GTGTTAATAT GCAACATCAT AAAGTGGCTA	4080
10	TTATcGGTGC CGGTGCTGCA GGTATAGGTA TGGCCATTAC CTTAAAAGAT TTCGGTATAA	4140
	CAGATGTCAT TATTTTAGAA AAAGGAACAG TAGGACATTC ATTTAAACAT TGGCCGAAAT	4200
	CGACCCGTAC GATCACGCCA TCATTTACGT CTAATGGATT TGGCATGCCT GATATGAATG	4260
15	CAATTTCCAT GGATACTTCA CCAGCATTTA CATTTAATGA AGAACATATT TCCGGAGAAA	4320
	CATATGCTGA ATATTTACAA GTGGTTGCCA ACCATTACGA GCTGAATATC TTTGAAAATA	4380
	CAGTTGTCAC AAATATATCT GTAGATGATG CATATTATAC GATTGCAACG ACAACAGAGA	4440
20	TATATCACGC GGATTATATC TTTGTCGCAA CAGGTGATTA TAATTTCCCT AAAAAgCCAT	4500
	TTAAATATGG TATTCATTAT AGTGAAATTG AAGACTTTGA TAACTTTAAT AAGGGGCaAT	4560
	ATGTGGTTAT CGGAGGTAAT GAAAGTGGCT TTGATGCTGC ATATCAACTT GCAAAAAATG	4620
25	GCTCTGACAT CGCACTTTAT ACTAGCACAA CCGGTTTAAA TGATCCGGAT GCTGATCCTA	4680
	GTGTTAGATT GTCACCTTAT ACACGTCAGC GACTAGGTAA TGTCATTAAG CAAGGTGCTC	4740
30	GCATCGAAAT GAATGTACAT TATACAGTTA AAGATATTGA TTTTAACAAT GGACAGTATC	4800
	ATATCAGTTT TGATAGCGGA CAAAGTGTGC TTACACCTCA TGAACCAATA CTAGCAACTG	4860
	GCTTTGATGC AACAAAAAT CCAATCGTTC AACAATTATT TGTGACAACA AATCAAGATA	4920
35	TTAAATTAAC AACACATGAT GAATCGACAC GTTATCCGAA TATTTTTATG ATTGGTGCAA	4980
	CAGTTGAAAA TGATAATGCC AAATTATGCT ATATCTATAA ATTTAGAGCG CGATTTGCAG	5040
	TACTTGCACA TCTTTTAACA CAGCGGGAAG GcTTACCAGC TAAACAAGAT GTCATTGAAA	5100
40	ATTATCAAAA AAATCAAATG TATTTAGATG ATTATTCATG TTGTGAAGTG TCATGCACAT	5160
	GTTAGAAGTG AAATATGATA TGAGAACTGG GCATTATACG CCCATACCTA ATGAACCTCA	5220
45	TTATTTGGTT ATTAGTCATG CGGATAAACT TACCGCAACA GAAAAAGCGA AATTAAGATT	5280
	ATTAATCATA AAACAGAAAT TAGATATTTT ATTGGCAGAA AGTGTAGTTT CTTcGCCTAT	5340
	AGCGAGTGAA CATGTGATAG AACCAATTGAC ACTATTTCAA CATGAGCGAC GACATTTAAG	5400
50	ACCTAAAATA AGTGCGACAT TTTTAGCCTG GTTGTGATA TTTTAAATGT TTGCATTGCC	5460
	AATCGGTATC GCTTATCAAT TTTAGATTG GTTTCAAAAT CAGTATGTGT CAGCATGGAT	5520
55	AGAATATTTA ACTCAAACAA CATTGCTCAA TCACGATATA TTACAGCATA TATTATTTGG	5580

ATTGATTAGT TTATCAACTG CTATAATTGA TCAAACAGGA CTCAAATCAT GGATGATATG 5700
 GGCAATTGAA CCGTCAATGT TATGGATAGG ATTACAAGGT AATGATATCG TGCCACTATT 5760
 5 AGAAGGGTTT GGATGTAATG CAGCAGCTAT TTCACAAGCA GCACACCAAT GCCATACCTG 5820
 CACGAAGACA CAGTGTATGA GTTTAATAAG CTTTGGTAGT TCTTGTAGTT ATCAAATAGG 5880
 10 TGCGACATTA TCTATTTTGA GTGTAGCTGG AAAGTCATGG CTATTTATGC CGTACTTAAT 5940
 ATTAGTACTT TTAGGTGGCA TCTTACATAA AGGATATGGT TGAAAAAGAA TGATCAACAA 6000
 CTTAGCGTTC CGCTACCTTA TGATAGGCAA TTACATATGC CAAATATACG TCAAATGTTG 6060
 15 CTACAAATGT GGCAAAATAT ACAAATGTTT ATCGTTCAAG CGCTACCTAT TTTTATCACA 6120
 ATCTGTCTTA TTGTTAGTAT TTTATCACTA ACGCCAATTT TGAATGTTTT ATCACAAATA 6180
 TTTACACCTA TATTATCGTT ATTAGGCATC TCGTCAGAAT TGTACCAGG GATTTTATTT 6240
 20 TCAATGATTC GAAAAGACGG CATGCTCTTG TTTAATTTGC ATCAGGGCGC CTTATTACAA 6300
 GGAATGACAG CAACACAGTT ACTACTACTT GTGTTTTTTA GTTCAACATT TACAGCGTGC 6360
 TCGGTCACAA TGACGATGCT TTTGAAACAT TTAGGTGGTC AGTCAGCACT AAAATTAATT 6420
 25 GGAAAGCAAA TGGTGACATC ATTGTCTTTA GTTATTGGTG TAGGCATCAT TGTAAAAATA 6480
 GTAATGCTGA TTATTTAAAA AAAATGAACT ATAAGTGAAT ATAGAGTCAT GTCAGTCAAT 6540
 30 AGGAGATCTA TCTTGAATA TGCTATTCAT ATGAAGTATA AGAGGAGAGT CGCAGATGAA 6600
 AATAGTTATT ATAGGTGGGT TTTTAGGTGG CGGTAAAACG ACTGTCTTAA ATCATTTGCT 6660
 CGCTGAATCA TTAAAGGAAT CGCTGAAACC AGCAGTCATC ATGAATGAAT TTGGGAAAAT 6720
 35 GAGTGTTGAT GGTGCCTTAG TATCTGAAGA CATACTTTA AGTGAAGTGA CAGAGGGGTG 6780
 TATCTGTTGT GCAATGAAAG CAGATGTATC AGAACAGTTA CATCAATTAT ATTTAAAAGA 6840
 GCAAACAGAC ATTGTATTGA TTGAATGTAG TGGGATTGCA GAACCGGTCT CTGTCTTAGA 6900
 40 TGCTTGTTTA ACGCCTATTT TAGCTCCGTT TACAACAATT ACACATATGA TTGGTGTAAT 6960
 AGACGCAAGC ATGTATAAAC ACATTAAATC ATTCCCTAAA GACATCCAAG GCTTATTTTA 7020
 TGAGCAATTA GCATATTGTT CTGTCTTATT TGTTAATAAA ATAGATTCAG CAGATGTTGA 7080
 45 AACAACGAGC AAATATTGA AAGATTTAGA AGTTATTAAAC CCAGAGGCCG ATATACAAGT 7140
 CGGTATGCAT GGCAGCGTCA CTTTGCCAAT ATCAGTTAGA CAAATGACAG CAACTTCTGA 7200
 50 CAATAAACAT AAGTCTTTAC ATCAAATGAT TAATCATCAA TTTGTGCAAT CACCAGTCAA 7260
 ATGTACTAAA GCAGAGTTTA TAAACGTTT AGCATGCCTT CCGTCTCATA TTTATAGGTT 7320
 GAAAGGGTTT ATGACATTTG AAGACACCGC ACATACGTAT CTCATTCAAT TTACACAAGG 7380

55

	CGGAAAGGGT ATTTCAAAAG AAGACTATCA ATGTTTGGAA CAGTAGTGTT TTCAGTGGAA	7500
	GAGAATGGTT AACATGCCTT CATGTATAAT AACGAGTTGA TTTGAACGTT TAAGCGTAAA	7560
5	TAAAAATAAG CTTGGTCAGC CATCAAATAT AATTTGAAAA CTGTCCAAGC TGTTTTATTA	7620
	GAGAACAATC AATTAACCCC ACATATTTAA TAATACATCA GCAAAGCCTT CAGGTTTTTG	7680
10	AATATAACCT AAGTGACCGC CTGGAATATC TACAATAGGT ATGCCAGTTT CTTTATTTAT	7740
	ATAAAAGTTA ACATCTGTG GGAAGGAGCC TCTAGAATCT GTCCCATTTA GTAGGGTGAT	7800
	TTTATCGCTG TATTTTGTGA AATCATCCAA AGTAATATCT GAATGCGTAT ATTGTCTAAT	7860
15	TTCAAATTCT GACCAGAACA TCGTACGTTT GACTGTCTCT ATACGTCCTT CTTCAGTATC	7920
	AGCAGGTTGA GACATCATTT TTGCATCAAT TGGTGCGATA TTTAATGTTT CGCCAAATGT	7980
	TTTCATGCCT TTTTCTAAGC CTTCTGTAA AATTTGATGC ACAATGTCAT CATTTTTATC	8040
20	TTTCCAATAA GTACTGTCTG GTAAAAATGT ATTAATTGGT GGTTTCGTGAA ATGCAATCTT	8100
	TTTAACGACT TCAGGGTAAT CTTTTAACAC ATGCATCGCA ACGATTGAAC CTGAACTTGA	8160
	ACCTAATATA TAGACAGGTT CATCACTTAA TGACTTTGCA AGTTCGGCAA TGTCTGTGC	8220
25	GTCGCGTTTG ACACGATAAT CACTGTCAGG GTTTGAAGCG GAATCAGGGA GTGGTTCAGT	8280
	TAACCTCGCTT TCTCCATAAT CACGACGATC AACGGCTACA ACAGTAAAAT GGTCTTTTAA	8340
30	CTGTTCTGCA AGAGGCAGAA AAATGTCTCC GGTACCGTTT GCACCAGGAA TAAAGATGAG	8400
	CACGGGTCCT TGTCGACTT GGTGGTATCG TAATTTAGCG CCTTGTAATT CTAAAGTTTC	8460
	CATATTCAAT GACCTCCATT TGTTAATTGT TAGGTGATAA ACCTAATAAT TTAGCACCAT	8520
35	TTGTATAACT TATTTCTCT TTTTCTTCAT CTGTTAAACC CAGTTCATCT AAAAATACAC	8580
	CTAATTTTTT AGGCTCAATA TATGGATAAT CAGCAGCATA AAGAATTCTA TCAATACCTA	8640
	CTTCTTTCTT GACTAAATCA AACTGTGGCT TCGTTAACAT GCCACTCGGT GTGATATAAA	8700
40	AATTATTTTT AAAGTAATAG CTTACAGGGT GGTTCAAATG TTCAGCGAAT AAAGCTTCAT	8760
	CCATACGTTT TAAGAAGAAT GGGATAAACT CACCCCAATG TCCAATAATC ATATTTAACT	8820
45	TTGGATAACG ATCAAAAATA CCAGATAATA CTAGATGTAT TGTATGAATG CCGACATCAA	8880
	TGTGCCAACC ATAACCAAAA CAAGCAAATG TTGCCGCAGT TACTTCAGGA TAATTTCCCTT	8940
	TATAGTATGA TTGATAAATG TCACTGTAA CTGGCGCGGG ATGTAGATAA ATCGGTACGT	9000
50	CTAAATTTTC AGCTGTTTTG AAAATAATGT CATATTTGTC TTGATCAAGA AAACCATCTT	9060
	GTGCACGTCC CATAATGAGC GCACCTTTGA ATCCTAAATC ATTGATGCAA CGTTCGAATT	9120
55	CTCGCGCTGC GGCTTCAGGC TCATTGATAG GTAAAGTTGC AAAGCCTACA AAGCGATTGG	9180

	TCTGACCAAC	CAAATTTGAA	GGAGAACCAT	TTCCATAAGA	TAAGACTTGA	ATTTGAACGT	9300
	CTTGATTATT	CATAAATTGG	ATACGTTTCAT	CATGATGTGA	TAATTCGTCG	GCATTGTGTA	9360
5	AACCTGTCTT	TTTTTcAAGG	CCTTCTAACA	TTACTTTTCAT	CGGTACACCT	TTAGGATCTG	9420
	CTGATATCGC	ATTCATCGTT	TCTTTTTGAA	TATCTTCAAT	GACATAATGT	TCTTCAAACG	9480
10	TAATACTTTT	CATTTACTTC	GCCTCCATAT	TGTATTGCAT	GTTTATTGCA	TCTATTGCAG	9540
	AAGCATTTTT	TATATACCTC	TAATTTCAAT	GTTTGTAACA	TAAAATTGAT	CTACCAAGGC	9600
	ATCTCTCCAT	CGCCATTAAT	AAATGTACCT	GTTGGGCCAT	CTGCACCAAT	CGTTGCTAAT	9660
15	TGAATGATTG	GCTTGATTCC	TTCAGAAACG	TGTTTGGAAT	TATTACTAAA	ATCACCAACT	9720
	AAATCAGTAT	TTGTAGCGCC	TGGATCAGCA	GCATTGATTT	GCATGTTAGG	TAATCCTTTA	9780
	GCGTATTGTA	GCGTTAGCAT	TGTTACTGCC	GATTTAGACG	AACAATAAGC	TAATGAATTC	9840
20	ACTTTAGATT	CAGCTGTTTC	GGGGTTTGTA	ACCATTCCAA	ATGAACCTAA	ACCACTTGAT	9900
	ACGTTGACGA	CAACAGGTTG	TTCAGATTTT	TCTAAGAGAG	GGACGAATGT	ATTCATCATT	9960
25	CGTACGATAC	CGAATACATT	CGTTTGATAT	ACTTCTTCAA	CGTCACGAGG	TGTCAATTTG	10020
	GAAGGTGCTG	AAAATTGACC	AGATATACCT	GCATTGTTAA	TGAGGATATC	AAGACGGCCT	10080
	TCTTTTTTCAG	CAATCATGTT	ATAAGCATT	TTGACTGAGT	AGTCACTTGT	AACATCTAAT	10140
30	TGTACATAAT	GAACACCTAA	TTTTTGATG	GCTTGTTGTC	CTCTTACATC	ATTCCGAGAA	10200
	CCTATATAAA	CTTTGTAACC	CAATGCTTTA	AGTGCCTCTG	CACTTGCATA	GCCTAACCCT	10260
	TTATTGCCTC	CTGTGATTAA	CACAATTTTA	GTCATTACGT	CCCACCTCAT	CTAAATAAAT	10320
35	GTTTAATAAA	TAATTTCTGT	ACGCTTCAAT	TGAAATATGG	CGATGCTCTA	TTTGGGAAGGC	10380
	AAATACACTA	GTTGATAATG	ATTGCAACAG	CATATCTGTT	TTGAAATCGT	GTAAGTGTCTG	10440
	TCATcGCTTT	TAAATAAGTC	ATAATAAAAA	TCAAATAATT	CTTGATAAAA	TGCGCTTTGG	10500
40	TAAAAACGTA	ATTTATTGTT	GCCTGCTTCA	ATACATTGCA	GTAGTGCCTT	ATTATCGATT	10560
	TTAAATTGTA	AAAGATAATC	TAACGACACT	TGCATAACCT	CATAATTAGA	ATGATAGTCA	10620
45	TCTTTAATTT	GCTTAAATG	AGTGATAAAA	ATATCAAGGT	CTCTTTGTAT	GACGTAGTAG	10680
	CATAAATCGC	TTTTATCTTT	GAAATGTCCG	TACAATGTCC	CCATACCGAT	ACCTAGTTCT	10740
	TTAGCAATAC	GATTCATACT	AATGTTTTCA	ACGCCTTCTT	CATCAAAAAG	TTTGTGCGCT	10800
50	ATTTCTTCAA	TTCGTTGCCT	ATTCTCTTTT	GCATCTTTTC	GCATGATTAC	ACCTACTTAA	10860
	AATTCTCTAA	AATTGACAAA	CGGATAACTC	TCCGTTTATT	ATAAAACGTG	TTAAGAAAGT	10920
55	TAGCAATGAA	TTTGCAATAA	CTATTAAATA	TCATAAAAGA	AAAGAGTGTT	GATAATGTCT	10980

ACCTTATCGG TTCAAATGAT TGCTGAAAAA CTGAATGTCA CTACAGAAGA TGTGGAAAAA 11100
 GTATTAGCTA TGACAGCGCC ACTAGGCATT TTTAGTCATC AATTACAACG ATTTATTTCAT 11160
 5 TTAGTATGGG ATGTCAGAGA TGTAATAAAC GACAATATTA AAGGAAATGG ACAAACACCA 11220
 GAACCATATA CGTATTTAAA AGGTGAAAAA GAGGACTATT GGTTTTTAAG A 11271

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6261 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

20 CAACCCGTTT AGAACAAAAT AAAAACCGTA CAATTTTATC ATCTTAATGA TTATTGTACG 60
 GAAAAACTTT TTTACATCAT ATCTGCATGT GCATAATCGA TATCGGTAAA TTTATTATAT 120
 TGTTCATAA AATGTAACCT AACTGTGCCT GTTGGACCGT TACGTTGCTT AGCAATGATA 180
 25 ATTTCAATTT CACCGTTTTC ATCATTCGTT TGTGGCTCGA AACCACCATC ATCGTCATCA 240
 TCTTCATCGC CGCCACGGTT ATAGTAATCA TCACGGTATA AGAATGCAAC GATATCGGCA 300
 TCTTGCTCAA TCGAACCAGA TTCACGAATA TCACTCATCA TTGGACGTTT ATCTTGTCGT 360
 30 TGTTC AACAC CACGAGATAA CTGACTTAAT GCGATAACTG GACATTTTAA TTCACGGGCT 420
 AATGCTTTTA ATGTACGAGA GATTTTCAGAA ACTTCCTGTT GTCTGTTATC GGACGCACGT 480
 35 GAACCACTAC CTTGAATCAA CTGTAAGTAG TCAATCACA TCATGTCTAA GCCATGTTCT 540
 TGCTTTAATC GACGACATTT AGAACGTAAA TCATTAATTC GAATACCCGG TGTATCATCA 600
 ATAAAAATCT TCGTACGTGA TAATTTACCT ACCGCTATAG TAAAACGACT CCAATCTTCC 660
 40 TCAGTCATAG TACCCGTTCT TAAGCGGTTT GAGTCAACAT TTCCAGAACT ACAAATCATA 720
 CGTGTGGCTA ACTGATCAGC ACCCATCTCT AGCGAGAAAA TACCAACTGT ATACATATCT 780
 TCATGCGTTG CAACTTTTGT TGCAATATTA AGTGCGAACG CAGTCTTACC TACAGATGGA 840
 45 CGCGCTGCAA GGATAATTAA ATCATTTTCG TTGAACCCTG CTGTCATTTG GTCTAAATCT 900
 CGATATCCTG TAGGTATACC TGGTGTGTTGA CCACTATTTT GATCAAGCTC TTCAGCTGTT 960
 50 TCATACACTT GTCCTAAGAC GTCTCGAATG TCTTTAAAGC CATCGCTTTC ACGAGAAGAT 1020
 GATAGCTCTA AAATTCGACG TTCTGCATCA CTTAAAATCG CATCTAGTTC AAGTTCATCA 1080
 TTATATCCAT CATTGGCAAT ACTATCTGCA GTTTGAATCA ATCTACGTTT TAATGCATGC 1140

	TCTGCAAGAT ATTGCGGGCC ACCCGCTTcA TTCAACGTAC CTTCCGTCGA TAATTGATCC	1260
	ATCAATGTTA CAACATCAAT TTCTTTATTA TCTTCATTTA AGTGCATCAT TGCACGGAAA	1320
5	ATATGTTGAT GGGCACCCCT ATAAAACGAC TCAGGAAGCA AAACCTCCTG AGTAGTATTA	1380
	ATCAATTCTG GATCTATAAT AATTGAACCT AAGACAGACT GTTCAGCTTC ATTGTTATGC	1440
10	GGCATTGAT TTTGCTCATA CATTCTATCC ATGAATGGTT ACACCTCTTA TTTCAATCCA	1500
	ACTTTATTGT TCAACTGTGT GTACGCGAAT TGTACCTTCA ACTTCTTTAT CTAATTTAAC	1560
	AGGTACATTC GTATATCCTA GGAATGAAT TCCATTGGT AAATCCATTT TACGTTTATC	1620
15	AATTTTAATA TCATGTTGTG CTTTGTAGTG TTCGGCAATT TGTTTGTAC TTAAGTACCC	1680
	AAACAATTTA CCACCTTCAC CAGTTTTTGC TGaTACTTCA ACTTCAATGT TTGATAACGT	1740
	TTCTTTTAAT GCITTAagCAT CTTCAATTTT TGTGTGGCGT TCTTGTMTTG CACGTTTTTT	1800
20	CTGTAACCTCT AATTGTTTAA GGTTACCTGG TGTGCTTCT ACAGCATAAT TCTTTTTCAA	1860
	TAAGAAGTTA TTTGCATAAC CTAAGTGTAC TTCTTTAACT TCACCTTTTT TACCTTTACC	1920
	TTTACCTTTA ACATCTGTG TAAAAATTAC TTTCATGCAT CTTCACTCCT ACTTAATTGT	1980
25	TCTGTAAATTG CTTGTTGTAA TTGTGCTATC GCCTCTTCGA CTGTACACACC TTTAAGTTGT	2040
	GTTGCCGCAT TGGTTAAATG TCCACCGCCA CCAAGTGCTT CCATTGTTAA CTGGACATTT	2100
30	ACTGAACCGA GTGAACGCGC AGATATACCA ATCAGATTAT CTTACGTCT CGCAACAACA	2160
	TATGATGCTT CAATACCTTC TAACTTAAAC AGTTCATCTG CTGCTGTGTC AACTGTTACT	2220
	GGATGATAAA TTTTATCGTC TGAACCATGC GcAATGGCTA TGCCATTATC TTCAACTTTT	2280
35	ACAGTTCGAA TTAATTCAGA TCGATTAATG TAAGTATCCA CATCATCTTT TAAGAAATGT	2340
	TGCGTTAAAA TCGTATCTGC ACCATGTGCA CGTAAATAAC TCGCTGCATC GAATGTTCTT	2400
	GATCCTGTTC GTAATGTAAA GTTTCTTGTA TCTACAATAA TACCTGCATA CATCACTGTT	2460
40	GATTCAAGAC GTGTTAAACG TTGTTCTGTT GGTGATATT CCAGTAACTC TGTTACCAAT	2520
	TCAGCTGTG AACTTGCGTA TGGTTCCATA TATATCAACA ATGGATTAGA GATGAAGCTT	2580
45	TCACCACGTC TATGATGATC GATAACAAC TACGGTTTG CTTTATTTAA GACATTTTCA	2640
	TCTAAAACCA GTTCCGGTTT ATGCGTATCA ACAATCACTA CGGTTGTCTT AGATGTCATC	2700
	ATATCCCAAG CATCATCTGA TGTAATAAAT CGCTCTCTTA ACTCTGGCTT TTTATCTATT	2760
50	TCGTTTCATCA CGCGTCGTAA TGTTGGATCA ATGTCAGTCT CATTAAATAC GATGTATGCT	2820
	TCTAAATTAT TCATCATTGC AAATCTAGAC ACACCGATTG CTGCACCAAT TGCATCTAAG	2880
	TCAGGACGTT TATGTCCCAT GATAATGACT TTGTCACCCT CTGCAAGGAT ATCTTTTAAC	2940

55

	CCATAGAAAC GCACATTACC ATTAATACTT TTAATTGCAA CTTGGTCGCC ACCGCGTCCT	3060
	AATGCTAAGT CTAGGCCTGA TTGTGATAAT TCACCTAAGT CGATTAAATT TTCAGTACCT	3120
5	TCACCAACAC CGATACTTAA TGTTAATTGG GCACGATAAC CAACACTTTT TTCACGTAAT	3180
	TGACTCAAGA TATCAAATTT AGATTCTTCT AAGTCAGCTA ATATTTTTTG ATTTAAATAG	3240
10	GCTACGAATT GATCGGAACT GTATCTTTTG AAAAATATAT TATACTCAGT TGCCCATCGA	3300
	CTAATGACAC GCGTTACCAT TGAGTTGATT TCCGAACGCT GCGTATCATT CATATTTTGC	3360
	GTAATCTCAT CGTAGTTATC TAAAAATAAT GTCGCAATGA TTGGTTTAGA ATTTTCATAT	3420
15	AGTTCATTG TTTGTACTTG TTCAGTTATA TCAAAGAAAT AGAGGCAGTG ATCATTCTCA	3480
	GAATAACGTA CTTGGAAATG ATACTGATTA TATCTATTT CAACGGATTT CACTCTATCT	3540
	AATGCTTTA AAATGTTTGG AAATACTTCA TTTACAGATT CAGAAATGAC ATTCGCTTCC	3600
20	ATATGATCTG TCATAAATTG GTTAACCCAT TCGATGTGAT CATTTTCATC TAAAACAATG	3660
	ATACCAATTG GTAAATGTTT GATTGCTTTA TTATTTGTTG TTGAAATTTG AGCACTCAAA	3720
	CCATCTACAT AACTATCCAT TTTCATTAAA GCTTGTCTGA ATAAAATGAT GCTAACAATA	3780
25	ATCATCACGA CAAGAACGAT AGATGCAATT AGTGCTATAA GACTATTAAA GATAAACCAT	3840
	ACACCCATTA AAACAATTGC TGTGATGATC ATGATGACAA ATGGTATTAG TAAAGCTTTC	3900
30	TTAGTGGACT GCCGATTCAT TATTCCACCT CTATTCACTT TTTAGAATTA TTTTTCATGA	3960
	TTGCTTCAA ATTCAAACCT AAATCGATAA CACCAAGTAG TCCTACAATA TGTGTCGTAG	4020
	GTGTCAGTAT TGTACCGATA ACCAATAGTA AAATCGTTAC TGCATTCGGC AAACCTTTCG	4080
35	CTTTACCAA GAAATGAATA ACACCTTAAAC CTTGAATATA CATTACTAAT GATAACACAA	4140
	GTTGGAAGTT TAAAAGAATG CTCTGGAACA CACTCGGTTG ACCTGTAAAT AATAAACATA	4200
	TGATAACAAT AATGTATATC CATAATAAAA TACCGCTCAT TTGCCACGCG AAAAGTGGCT	4260
40	TAAATACAGG TGTAGCGATT TTAAATTTTC GTAAAATCGG AAATGTAAACG ATTAAGTTAA	4320
	TTAAGACGAT TAAAAATGTA ATGATAATGA TGAAACCTGG TAATTGAACG GTCGCTTGTC	4380
	TAAACCCCTT TTCTAATATT TGGGTCATAT TCGCATCGGC ACCGCTCATC GTAATCGCTT	4440
45	CATGTAATGT TTGCTTGAAA GGTTTTACTA TGCTCGCTGA TGGTGGAAATC CTTCCGAATG	4500
	TTTGTAGTAA CATAAAAGCG ATTAATGAAA TTAATCTCAT CGCTACTGTT GTTACGTATA	4560
50	ATATTCTTTC TTTAGACGTT CTTTCTTTGA GCAATTGACC AATAATTAAA CTTGCAATTA	4620
	AGACTAATAT GATGGCACTT AAAACGAAAG TATTACCTAA AACAGTTGTT ATAATTACTG	4680
	TAATAAGTGC ACTAATCCCG AAAGATTGTA TTGATTTATT CCATAAAACG ATACCTGGTA	4740

55

CAAATACCAA CGCAATCGTT GCAATTATTG TTGCTTTAGG TTGTATTTT GAAAACACAT 4860
 AAGCCACTCC CATATTTTTA ACTATAGCTA TTATTTTAAC CTCTTTAATG AAAATTAACA 4920
 5 ATTTATAGAT TGTATGCTTC TATTTCAATT AATTGAATAA TAACTTTCAT GTTTTATAAG 4980
 TAATTAACAT ACTCATTGGA ATCGCTTTTG TGTGCTTTCA TTTTCAACAT GATTATTTAA 5040
 TCCCACTACA TAGCAATCAA GCTTGATTGA GATTTACAAT ACATTTCAC TCTCATGTAC 5100
 10 TCTAGATGTT TTTGAATATG ATAAGTGTGA TTTAGTGGCT TCATTCTTTG AAAATATATA 5160
 TTATTACTTA CGCTTAAAAT GCTTTAAATT TAAGAAATGA TATAAGTTAG GTGCCCAGGT 5220
 ACTAAAGTTT AGTAGGATC CATCATGCCC AACATTATCA GGCACGAAGA AATGACGATG 5280
 15 ATATTTAAAA CGTTCACCTA ATGCACGAAC TTGATCATCC GGATATAGCA AATCATCTAT 5340
 GAACCCCATC GTTAACACTT TTGTTTCTAA ATTTTAAAA ACATGCGTTA CGTCTGTGCG 5400
 20 ACCTCGGTCA ATGTTGTGAC TATCCAATAC ATCTAGCAGT GTCAGATAAC AATTCAAATC 5460
 AAAATGTTCT TTAAATTTAT TACCTTGATG TTGTTGGTAT GCGACTACTT CATCCGGCGT 5520
 AAAACGTTCA TCATAACTTT TTGATGATCG ATATGTCAAA AAACCTAATT GCGTGCAAT 5580
 25 ACTTAGACCT TCCTTACCAC CAAGATGAAT GGCTGCGCTT GCAATTCAT TGAAAGCTCT 5640
 ACTATAAGAT GATGTTGAC TTGTTGCAGC AAGGATAATG GCTTTATCTA CTTCAAACCTG 5700
 TTGATTGTAG AGTAGTTCCA TTGCTTGCAT ACCTCCAAGA CTTCCCCCTA TTAAAATATT 5760
 30 AATCTTATCA TAACCAAGGG CTGTGATACC TCGTTCATTC GCTCTGACTA TATCTCTTAA 5820
 TGTTAATTTT TTAGGAAAAT GAGGGTCGTT TAAAGGTGAA CTTGAACCGA AAGGACTACC 5880
 AATAACATCA AATGTTAAAA ATTGATAATC GTGAATGGGT ATATATCCCC CATCAATAAT 5940
 35 TTCTCGCCAC CAACCCGGAT AATCATCTGT TCCATATGTT AAATGATTGC CAGTTAATGC 6000
 ATGACAAACT ACAACTAATG GTTGTCCATG ATAACCGACA TGCTCATATC TCAAACGCAA 6060
 40 GTTATCTATG ACTTCCCCAG ATTCTGTAAT AAATCCCCT AAATTTAAAG TATCTACTGT 6120
 GTAATTTGTC ATTGTTCTTT CTCCTTAAA CAAAAAACT TCTCACCCTA TTGAAAAGTA 6180
 AGAAGTCTTT ATACTTATCA TTCGAGTAAC TCGTTGGTTT TAGCACCGTG CTATAAAGTC 6240
 45 GGTGCTGAA GTATCACAGG G 6261

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1222 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

5 ATGCGATTAA CTCTGGAAAT ATCTTTTCCA TATTTACGTh TTAAATTATT CAGCAAATTC 60
 ATACGAGaTT CATACTCGTT yAACACTTGT TCGTCGAATT CTGTATTAGC CATTTTCATCA 120
 TATAACTCAT GTTTTGCATC TTCTAAAATG TAGTAAAATT GATCAATATC TTCTTTTAAT 180
 10 TTGTCATATT TGTTTGAAC TATATCGTTT ATTGTTAACA AATGGTTGCT TAGTTCATAT 240
 AAACGATCAG TGATAGCATT TTCATCCGTT AATGTCATAT ATGCGTTATT AAGCGCTAAG 300
 CTTAATTTTT CAGAGTTTTG AATGCGTTTA ATATCTATTT CAAGTTGCTC TATTTGCGCT 360
 15 TCTTTTAGAT GTGCTTCAGA CAATCTTCT AATTGGAATT TCATTAAATC TAAACGCTGT 420
 AGCAATGCTT GGTCTGCTGA TTCTAAATCT TCTAACTCTT GCTTTTTGGC TTTATAATTT 480
 TGAAAAGTTT GGTGATATTT ATCCAACAAA TCTTGATAAC GTGATTCTGC GTAATTATCC 540
 20 AATAATGTTA AATGGTATTT TTGTTTCAAC AAAGACTGCG TTTTCATGTTG GCCATGAATA 600
 TCTAATAATT CTTGCATAAC TTTTCGTAAA TCTTGTAAG TAACTGTTTG ATTATTAATT 660
 TTACAAAGAC TTTTACCAGA GCTGAAAATT TCCCGTTTAA CTAATAAAAA ATCTTCATCT 720
 25 ACATCAATAT CCATATTTTT CAATATATGT ATAGCATCTT TACTCTCGTC AATATCAAAT 780
 ATACCTTCGA TGACAGCCTT TTTTCACCA TGTCTTACAA AATCAGATGA AGCTCTCATT 840
 CCAATTAATT GTCCAATTGC ATCTATAATA ATTGACTTAC CTGAACCCGT TTCACCACTT 900
 30 AAAACAGTTA AACCATCAGA AAATTGAATT TCTAATFCTT CAATAATAGC AAATTGCTTG 960
 ATTGATAAGG TTTGTAACAT AAACATCATG CATCCTTATA ACAAATTGAA AATTCTTGAC 1020
 35 TTGATTTTCAT CACTTGCCCTC TTTGCTTCGA CAAATAATTA AACAAGTATC ATCACCACAA 1080
 ATTGTCCTA GTACTTCTTC CCAATTGATT TGGTCTAATA TAGCTCCAAT AGATTGTGCA 1140
 TTACeAGGTA TGTTTTTAGA ACAAGTAAAT TATCAGTACC ATCTATATTA ACAAAGGAAT 1200
 40 CCATTAAATA ACGTCCCAAT TT 1222

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTTGTTATTA TTACnTnAAA TAATTGCATT ACTTTTTACT GATGGTACAA CTTTCCATCC 60

TTCTTTTGGC ACGACATAAT TGTCTTTATC TTGAACTAAA TATCCGCCAG ATACTGAAAC 180
 AAACTCTTCT TCGTTACTGT CTATAGTCAT ATCAATTTCT AATAATCTTA CATTCTTCTT 240
 5 TTGTTTTAAA ATATCTAATG CTTTCATCTGT AAATTTTGGT GCAATAATGA CTTCCAAAAA 300
 GATACTATGC AATTGCTCTG CTAAGTCAGG TGTTACAGCT CGGTTTAATG CAACAATTCC 360
 ACCAAATATT GATTGACTAT CCGCTTCATA CGCATGTTGA AATGCTTGTT CTATCGTGTC 420
 10 ACCGATACCA ACACCACATG GATTCATGTG TTTAACCGCA ACTGTAGCAG GTGTATCAAA 480
 CTTTTTAACT AAAGCTAGTG TAGCATCTGC ATCTTTAATA TTGTTATAGC TTAATTGTTT 540
 CCCATGTAAT TGTTTAGCGC CTGCAATCGT GTGCTTAGCA TTCGAAGTTC TCACAAAATA 600
 CGCTGATTGT TGTGGATTTT CTCCATATCT TAAAGTTTCT TTATCCCCTT TAAAGAAACG 660
 TACAATCGCT TCATCATATT CTGCAGTATG CTCAAAAAC TTAATCATTA ATGATTGTCT 720
 20 ATATGACTCA TCTAACGAAT CGTTTCTTAA TCGCGTCAAT ACTTCTTGAT AATCTGCCCG 780
 ATGTACAATT GTTGTTACAT GTTTATAGTT TTTAGCTGCA GCACGTAACA TTGTTGGACC 840
 ACCAATATCA ATATTTTCAA TTGCTTCGTC CATCGTCACA TCAGGGTTTG CAACAGTTTG 900
 25 TTGGAATGGA TATAAATTAA CTACTACCAT ATCAATTAAA TCTATATGTT GTTCTGATAA 960
 TTCATTTAAA TGCTGCGGTT TATTTGATC AGCTAAAATG CCACCATGAA CAGCCGGATG 1020
 T 1021

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3759 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCATTCACTC CTAAATTGTT ATTACACTAT TACACaTAGC TAATCATCAA TGTGAAATCA 60
 CCTTCAAAGA CACTATCCAA ATCTTCAGAA GTCAAAATAA AGTTTGTTACC AGTAGTCAGT 120
 45 TTGAAAATTT CACCATCGAC AATCATTTGC CCTTCGCCTT CCAACACTGT AACTAAACAG 180
 AACTCTCTAG GCTTCATATA ATTTAACGTG CCAGAAATTT CCCATTTAAC CAATGTAAAG 240
 AAATCATTCG ATACAATGTG TGTACACTTA TGGTTTTCAA TAATTTGCTT TTCAGGCAAA 300
 50 ATATTAGGTA ATGGTGCATT GTACTGAATA ACGTCTAAAG CTTTTTCAAT ATTTAACGGT 360
 CTATCATTAT ATTGATTATC TTGACGATTG AAATCATAAA GTCTATATGT AATGTCTGAC 420

	ATAAAAaAGa	ATTcYCCAGG	kTTTACtTTA	AtatATCyAA	gTatCGaCtC	tATCGTTCCG	540
	TGTTGAACAT	GATTcGCAAC	TTCTTCTCTA	GACTCTGCTA	ATGTCCctAT	AACTATTTCT	600
5	GCATCTTCTT	CTGCATCTAT	AATATACCAA	CATTcAGATT	TGCCATATTG	CCCgTTTTCA	660
	TGCTCATAAG	CATAAGAATT	ATCAGGGTGC	ACATGAATAG	AAAGTGATTc	TCTTGcATCC	720
	ACTATTTTAg	TTAGAAGCGG	AAAATCTTTG	CTTGGGAAAT	CACCAAACAA	TTCACGATGT	780
10	TCTGACCAAA	TACGGTCTAA	TGTTTGACCT	TGATATGGTC	CATTAATAAT	CTCGCTCGTA	840
	CCATTTGGAT	GTGCTGACAC	ACACCAACAT	TCCCCAGTT	GTATCATTGT	CTAATTGATA	900
15	TCCAAACTCA	CTTAGACGTT	GACCGCCCCA	TAATTTTGTT	TTAAAAATTG	GTTGTAAAAA	960
	TAATGGCATT	GTTGCACCTC	CATTGTGATT	AAGTAAGCAA	TAGAACTCTG	ATGTTGTTGT	1020
	TCCATTATAT	TTTGATTTTG	TTCTCATTTA	CATCGTATTA	TTAACTTCCA	CATTTCAAAT	1080
20	TAACTATTAG	TGATTGTACC	ATATTTACTA	ACATTGCAGT	ACTGCCAATT	AAAAGnGCTT	1140
	CACTTAAATT	TACAGTACTT	TAACATTTTC	AAAAATTTAT	AGCATAGAGA	TTATATCTCT	1200
	CTTACATTTG	TACATATTTT	CCTTTAAATT	TACTCGCCCC	TTATACCAAT	TAATAaACAA	1260
25	CTTTAATAGT	TGTGCCATAC	ATTGTTCAAA	TTCTTTGTAA	AACGCATAGA	CAATACGTAC	1320
	TTATTCATAC	TTATAATTCA	TCATTTTCAA	AAAATAACGA	GTTACGAAAA	AGTAACCCGC	1380
	TTCAAATCAT	ATTTACTATC	CTTATTAAAT	CGTTTTCATTT	TCAAATTGAG	TTAAAGCATC	1440
30	TTTAATGTCC	TGATCACCAC	TAATAATTG	AAACTCTTGG	TGATTAAAAAT	GATTGGATGT	1500
	GACAATTTCT	TTTAATACTG	TCGCAACATC	TTCTCTAGGA	ATTTACCTT	TACCATCAAA	1560
35	ATATTGTGCA	GCTTCTATCT	TTCCAGATCC	TGCTGCATTT	GTAAGTGCCC	CTGGATGTAA	1620
	AATTGTATAA	TTCAAACCTG	nAACGTCTTA	AATAGTCATC	AGCGTAATGT	TTAGCTATTG	1680
	TATATGGCTT	TAAATCACCG	CTATCATCAA	AAGCCTGACG	TCTCGAATCA	TATGTTGAAA	1740
40	CCATGACATA	GTGTTTAATA	TTGGCCTCTT	TACTCGCAAT	CATTGATTTA	ACAGCACCAT	1800
	CTAAATCGAC	AATAATTGTT	TTATCTGCAC	CCGTGTTCCC	TCCAGAACCT	ACTGAAAAGA	1860
	TAACTTTATC	GAATGGTTTA	AACGTCTCAG	TTAAAGTCTC	TATTGAATCA	TTTTCAACAT	1920
45	CAACAAGAAT	TGCTTTCATA	CCTTGcGATT	TTAACGCATT	AAGTTGATCT	GATTGCCTAA	1980
	CACCAGCAGT	AAATGGTACA	TTTTCTTTTG	CTAATTGTTG	CACTAGTAAC	GAACCTACAC	2040
50	CGCCATTAGC	ACCTATAACC	AAAATATTCA	TTTACAACAC	TCTCCTATkT	ATTATTCTCT	2100
	ATGCCATACC	ACTTTATGAG	ATATGTAAAA	CTTGTTACAA	CTATAAAAAT	CAATTGACAT	2160
	ACTACTGGGA	ACGTATTAAA	TTAATATATG	AACAAATATT	CATATGAAAG	GATTGTcATA	2220

55

tCaAGGCATT AGcGATTACA ATCGAATACG TATCaTGGAA TTGTTATCaG TCAGCGAAgC 2340
 AAGTGTGGT CACATTtCAC ATCAATTGAA TTTATCTCAA TCAAATGTCT CGCACCAATT 2400
 5 AAAATTACTT AAAAGTGTGC ATCTTGTGAA AGCAAAACGA CAAGGCCAAT CAATGATTTA 2460
 TTCATTAGAT GACATCCACG TAGCAACTAT GTTAAAGCAA GCCATACATC ACGCGAATCA 2520
 10 TCCTAAAGAA AGTGGGTAT AATATGTCTC ATTCACATCA TCATCATGAC CATATGCATA 2580
 GTCATGTAAC TACAAATAAT AAGAAAGTAT TGTTTATATC GTTTTAAATA ATCGGTCTAT 2640
 ATATGTTTAT CGAAATCATC GGCGGTCTCC TTGCTAACAG CTTGGCATTa CTATCTGACG 2700
 15 GTATCCATAT GTTTAGCGAC ACATTCTCAT TAGGTGTTGC ACTTGTCGCA TTTATTTATG 2760
 CTGAAAAGAA TGCCACAACt ACAAAAACAT TTGGTTATAA ACGTTTCGAA GTACTCGCAG 2820
 CGTTATTTAA CGGTGTAACG CTTTTGTAA TAAGTATTTT GATTGTTTTT GAAGCGATTa 2880
 20 AACGTTTCTT TGTTCCTTCT GAAGTTCAAT CAAAAGAAAT GTTAATCATT AGTATTATCG 2940
 GTTTAATTGT CAATATCGTT GTTGCATTCT TTATGTTTAA AGGCGGCGAC ACTTCACACA 3000
 ATTTAAATAT GCGTGGTGCT TTTCTACATG TTATCGGAGA CTTATTAGGT TCAGTTGGCG 3060
 25 CCATTACTGC AGCTAKTTTA ATTTGGGCAT TTGGATGGAC AATCGCCGAT CCTATCGCAA 3120
 GTATTTTAGT TTCCGTTATT ATTTTAAAAA GTGCTTGGGG TATCACAAAA TCTTCAATTA 3180
 ACATTTTAAAT GGaAGGCACA CCAAGTGATG TTGATATAGA TGAAGTTATA ACTACTATTA 3240
 30 AAAAGGATTC ACGAATACAA AGTGTGCATG ATTGCCATGT TTGGACAATT TCAAATGATA 3300
 TGAATGCATT GAGTTGTCAT GTTGTGTAG ACCATACATT GACAATGAAA GAATGTGAAT 3360
 35 TATTATTAGA AAaCATTGAG CATGATTTAT TACATTTAAA TATTCACCAT ATGACTATTc 3420
 AATTAGAAAC GCCTAATCAC AAACATGATG AATCGATTAT ATGTTcAGGA ACACATAGTC 3480
 ATTCaCATAA CCATCATGCT CATCATCACG CGCATGTACA TTAATAATTT TAACCTACTG 3540
 40 CCATTGCATC GATTAAACTT TTCAATGGCA GTAGGTTTTT TATGTCTTTA TGGCGACTTG 3600
 TTTGGTCTTT GATGATGCAA TGTTTATTAA CAAATTTTCA ACTATTATTT CTTACATTAG 3660
 TCATATTTTT GACAATTTAC TATTATAAAT CTCTAACTTT AGTCACTTTA ATTAATTTTT 3720
 45 ATTAGATATT AATATGAAAA TAACGTGTTT TTGTTATT 3759

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 13086 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

	TAATTATCGC GCATAACAAA ACATTAGCAG GACAATTATA TAGTGAGTTT AAAGAATTTT	60
5	TTCCTGAAAA CAGGGTGGAA TACTTTGTAA GTtACTATGA TTATTATCAn CCAGAGGCAT	120
	ACGTACCGTC TACTGACACT TTTATTGAAA nAGATGCCTC AATCAnTGAT GAAATTGATC	180
	AACTACGACA TTCTGCTACA AGTGCATTAT TTGAACGCGA TGATGTAATT ATTATTGCTA	240
10	GTGTAAGTTG TATATATGGT TTAGGTAATC CTGAAGAATA TAAAGATTTA GTAGTAAGTG	300
	TTCGAGTTGG TATGGAAATG GATAGAAGTG AATTACTTAG AAAACTTGTC AGATGTGCAA	360
	TATACACGAA ATGACATCgA TTTcCAACGA GGAACGTTTC GAGTGCGTGG TGATGTAGTG	420
15	GAAATATTCC CAGCCTCTAA AGAAGAACTT TGTATAAGGG TTGAGTTTTT CGGCGATGAG	480
	ATTGACCGTA TCCGAGAAGT TAACTACCTA ACAGGTGAAG TGTTGAAAGA AAGAGAACAT	540
20	TTTGCGATAT TCCCAGCTTC TCACTTCGTA ACACGTGAAG AAAAGTTGAA AGTTGCGATT	600
	GAACGTATTG AAAAAGAATT GGAAGAACGA TTGAAAGAAT TACGAGATGA GAATAAATTA	660
	CTAGAAGCGC AAAGGTTAGA ACAGCGTACC AACTATGATT TAGAAATGAT GCGAGAGATG	720
25	GGATTCTGTT CAGGAATTGA AACTATTCC GTACATTAA CTTTGCAGCC ACTGGGTTCC	780
	ACACCATATA CTTTATTGGA TTACTTTGGC GATGATTGGT TAGTAATGAT TGATGAATCA	840
	CATGTGACAT TACCGCAAGT TCGAGGCATG TATAACGGAG ACAGAGCGCG TAAACAAGTT	900
30	TTGGTGATC ATGGGTTTAG ATTACCGAGT GCATTAGATA ACCGTCCACT TAAATTGAA	960
	GAATTTGAAG mAAAGACAAA ACAACTTG TG TATGTATCTG CAACGCCTGG ACCATACGAA	1020
	ATTGAACATA CGGATAAGAT GGTGAACAA ATTATTCGTC CTACTGGTTT ACTGGATCCT	1080
35	AAGATTGAGG TTAGACCTAC TGAAAATCAA ATTGACGATT TATTAAGTGA AATTCAAACA	1140
	AGAGTgAGCG TAATGAACGC GTACTTGTTA CAACGCTCAC TAAAAAGATG AGTGAAGATT	1200
	aACCACATAC ATGAAAGAAg CGGGTATTAA aGTLAATTAT CTGCATTCAG AAATCAAGAC	1260
40	ATTAGAACGA ATTGAAATAA TTAGAGACTT ACGAATGGGT ACATATGATG TTATCGTAGG	1320
	TATTAATTTA TTAAGAGAGG GTATTGATAT ACCAGAAGTT TCTCTAGTTG TCATATTAGA	1380
45	TGCAGATAAA GAAGGGTTTT TACGTTCTAA CCGCTCATTa ATTCAAaCAA TAGGTAGAgC	1440
	TGCGCGTAAC GATAAaGGTG AAGTCATTAT GTATGCCGAT AAAATGACTG ATTCGATGAA	1500
	GTATGCAATT GATGAGACAC AACGTCGTCG AGAAATACAG ATGAAACATA ATGAAAAACA	1560
50	TGGTATTACA CCTAAAACAA TTAATAAAAA AATACATGAT TTAATTAGTG CTACTGTTGA	1620
	AAATGACGAA AATAATGACA AAGCACAAAC TGTGATACCT AAGAAGATGA CGAAAAAGA	1680

55

	TTTCGAGAAA GCTACAGAAT TAAGAGATAT GTTATTTGAA TTAAAAGCAG AAGGGTGACA	1800
	AGTAAATGAA AGAACCATCC ATAGTAGTAA AAGGTGCTCG TGCGCATAAC TTGAAAGATA	1860
5	TTGATATCGA ACTACCTAAA AaTAAATTAA TTGTTATGAC AGGTTTATCT GGGTCAGGTA	1920
	AATCGTCATT AGCATTTCGAT ACTATATATG CTGAAGGACA ACGACGTTAT GTTGAATCAT	1980
	TAAGTGCCTA TGCGCGTCAA TTTTtagGCC AAATGGACAA ACCAGATGTT GATACAATTG	2040
10	AAGGATTATC GCCAGCAATT TCAATAGATC AAAAAACAAC AAGTAAAAAT CCAAGATCAA	2100
	CTGTAGCAAC AGTAACAGAA ATATATGATT ATATACGTTT GTTATATGCA CGTGTTGGTA	2160
	AACCTTACTG TCCAAATCAC AATATAGAAA TTGAATCGCA AACAGTACAA CAAATGGTTG	2220
15	ACCGCATTAT GGAATTAGAG GCACGTACAA AGATTCAATT ATTAGCACCT GTCATCGCTC	2280
	ATCGTAAAGG TAGTCATGAA AAGCTAATCG AAGATATTGG TAAAAAAGGT TATGTACGTT	2340
20	TAAGAATCGA TGGCGAAATT GTTGATGTAA ATGATGTACC TACTTTAGAT AAGAACAAGA	2400
	ATCATACAAT AGAAGTTGTT GTAGACCGAT TAGTTGTAA AGATGGAATT GAAACACGAC	2460
	TAGCTGACTC TATAGAAACT GCCTTAGAGC TTTCAGAAGG ACAATTAACA GTCGATGTCA	2520
25	TTGACGGGGA AGACCTTAAG TTTTCAGAAA GCCATGCTTG TCCTATATGT GGATTTTCAA	2580
	TCGGAGAGTT AGAACCAAGA ATGTTTAGCT TTAACAGTCC TTTTGGTGCT TGTCCGACAT	2640
	GTGATGGCTT AGGCCAAAAG TTAACAGTCG ATGTAGACTT GGTTGTTCCC GACAAAGATA	2700
30	AGACGCTAAA CGAAGGTGCA ATAGAACCTT GGATACCGAC GAGTTCTGAT TTTTATCCAA	2760
	CATTGTAAA ACGTGTGTTG GAAGTTTATA AAATCAATAT GGATAAACCT TTTAAAAAGT	2820
	TAACAGAACG TCAACGTGAT ATTTTATTGT ATGGTTCTGG TGACAAAGAA ATTGAATTTA	2880
35	CATTACACA ACGTCAAGGT GGTACTAGAA AACGAACAAT GGTTTTCGAG GGTGTAGTTC	2940
	CTAAATATAAG TAGACGATTC CATGAATCTC CTTCAGAATA TACACGTGAA ATGATGAGTA	3000
40	AATATATGAC TGAACCTTACCT TGCGAAACTT GTCATGGAAA GCGATTGAGT CGTGAAGCKT	3060
	TATCTGTTTA TGtagGTGGT TTAAATATTG GTGAAGTAGT CGAATATTCA ATCAGTCAAG	3120
	CGCTGAACTA TTATAAAAAC ATTGATTTGT CAGAACAAGA TCAAGCGATT GCAAATCAAA	3180
45	TATTGAAAGA AATTATTTCC CGACTCACTT TTTTAAATAA TGTGGGACTT GAATATTTAA	3240
	CGTTAAACAG AGCTTCAGGT ACACTTTCAG GTGGTGAAGC ACAACGTATT CGATTAGCAA	3300
	CGCAAATTGG GTCGCGTTTG ACTGGTGTCT TATATGTATT AGATGAGCCA TCAATTGGAC	3360
50	TGCATCAAAG AGATAATGAT CGATTAATTA ATACACTTAA AGAAATGAGA GATTTAGGAA	3420
	ATACTTTAAT TGtagTTGAA CACGATGATG ATACAATGCG TGCGGCTGAT TACTTAGTGG	3480

55

	AGGTAATGAA AGATAAAAAA TCATTAACAG GACAATACTT GAGTGGTAAG AAACGTATTG	3600
	AAGTACCTGA ATATCGCAGA CCGGCTTCAG ATCGTAAAT TTCTATACGT GGAGCTAGAA	3660
5	GCAACAATCT TAAAGGGGTT GATGTGGACA TACCACTATC AATCATGACG GTTGTTACAG	3720
	GTGTATCAGG TTCTGGTAAA AGCTCATTAG TAAATGAAGT ATTATACAAA TCATTAGCTC	3780
	AAAAAATTAA TAAATCTAAA GTAAAGCCAG GATTGTACGA TAAGATTGAA GGTATTGATC	3840
10	AACTTGATAA AATTATTGAT ATTGATCAAT CACCAATAGG TAGAACGCCA CGCTCTAATC	3900
	CAGCAACATA TACTGGTGTG TTTGATGATA TACGTGATGT GTTTGCGCAA ACAAATGAAG	3960
	CTAAAAATTCG AGGATATCAA AAAGGGCGTT TTAGTTTTAA TGTAAGAGGT GGACGCTGTG	4020
15	AAgcTTGTAA AGGTGACGGT ATTATTAAAA TTGAAATGCA TTTTTTACCT GATGTTTATG	4080
	TTCTTGTGA AGTGTGTGAT GGTAAACGAT ATAATCGTGA GACACTAGAG GTTACTTACA	4140
20	AAGGTAAAAA TATTGCTGAC ATTTTAGAAA TGACTGTTGA AGAAGCAACA CAATTTTTTG	4200
	AAAATATTCC TAAGATTAAG CGCAAGTTAC AAACACTAGT TGATGTTGGT CTTGGATACG	4260
	TCACATTAGG TCAACAAGCT ACAACGTTAT CAGGTGGTGA GGCTCAACGT GTGAaACTTG	4320
25	CATCTGAAC TCAATAACGT TCAACTGGTA AATCTATTTA TATCCTAGAT GAACCGACAA	4380
	CAGGGTTACA TGTGACGAT ATTAGTAGAT TATTAAAAGT ATTAAACCGA TTAGTTGAAA	4440
	ATGGTGATAC TGTGTAAATT ATTGAACATA ACCTAGATGT TATCAAAACA GCAGACTATA	4500
30	TTATAGACTT AGGTCCTGAA GGTGGTAGTG GCGGTGGTAC TATTGTTGCG ACTGGCACAC	4560
	CCGAAGATAT TGCTCAGACA AAGTCATCAT ATACAGGAAA GTATTTAAAA GAAGTACTTG	4620
	AACGAGATAA ACAAATACT GAAGATAAAT AAGATTAAAA GAAGTGAAGG ATGTTATAAA	4680
35	TTTATCCTTC GCTTCTTTTT ATTAATTTAG TAATGAATAG TAGAAAGAAA AGATGCGTAA	4740
	AAAGaATTAT GTTAAGATAG GGTCAATCTA GAGTAGTTAA ACATAAATCG AACTGGGAGT	4800
	GGGACAGAAA TGATAAAGAA TCACTAATGA TTTATTATGT AGTGGTTCTT TGTCATTAGC	4860
40	CACAGCTATT GTGTACTTAA AAATAGGaT GCaTgAGTGC AACTCATGCA TAAGaAATAC	4920
	TAATTTCTAA AGAAAAAGTA TTTCTTTATG TTGGGGCCCC GCCAACTTGC ATTGTTTGTA	4980
45	GAATTTCTTT TCGAAATTCT TTATGTTGGG GCCCCGCCAA CTTGCATTGT TTGTAGAATT	5040
	TCTTTTCGAA ATTCTTTATG TTGGGGCCCC GCCAACTAAT TCCAATATAT CATTGTAGAG	5100
	CTTAGGTCAT TGATTTTTTG CTCGGACTTT TATGGCGATA TGAACCATGT AAATTAAGCA	5160
50	AGCAATAAAT TAATGATTGA TATTGACTTG TAAATAATA ACAATAATGA ACAATTAATA	5220
	TTTATTTTAG CTTTCAATG TAGATTGGTG TTATATTTTT GATATGATAA GAAGAGATGT	5280

55

ACATTAAAGT TAGATTTAAT CGCTGGTGAA GAAGGACTAT CGAAGCCAAT TAAAAATGCT 5400
 GATATATCAA GACCGGGCTT AGAGATGGCA GGTATTTTTT CACATTATGC GTCAGATAGA 5460
 5 ATACAACTAT TAGGAACAAC GGAACATCG TTTTACAATT TATTACCAGA TAAGGATCGC 5520
 GCAGGTCGTA TCGTAAACT ATGCAGACCA GAAACGCCTG CAATTATTGT GACACGTGGA 5580
 TTGCAGCCAC CAGAAGAATT AGTTGAAGCT GCAAAGAAT TAAATACCCC ACTTATAGTT 5640
 10 GCTAAAGATG CGACTACAAG TTTAATGAGT CGCTTAACAA CGTTTTTAGA GCATGCACTT 5700
 GCAAAGACGA CATCTTTACA TGGTGTTTTA GTAGATGTTT ACGGTGTTGG TGTACTAATT 5760
 ACCGGTGATT CAGGAATAGG TAAAAGTGAG ACTGCGTTGG AATTAGTTAA ACGTGGGCAT 5820
 15 AGATTAGTAG CAGATGATAA TGTAAGAAATA CGTCAAATTA ATAAAGATGA ACTAATAGGG 5880
 AAACCACCAA AGTTAATAGA ACATCTATTA GAAATACGTG GACTAGGTAT TATCAATGTT 5940
 20 ATGACTTTAT TTGGCGCGGG TTCAATATTA ACTGAAAAAC GAATTAGATT AAATATTAAT 6000
 TTGGAAACT GGAACAAGCA AAAGTTATAT GACCGCGTAG GTCTTAATGA AGAGACGCTA 6060
 AGTATTTTAG ATACTGAAAT CACTAAAAAA ACAATACCTG TAAGACCTGG TAGAAATGTT 6120
 25 GCGGTAATTA TTGAGGTCGC TGCAATGAAC TATCGATTAA ATATCATGGG CATTAAACAG 6180
 GCCGAAGAAT TTAGTGAAAG ATTAAATGAA GAAATTATCA AGAACAGTCA TAAGAGTGAG 6240
 GAGTAGGTTG AATGGGTATT GTATTTAACT ATATAGATCC TGTGGCATT TAACTTAGGAC 6300
 30 CACTGAGTGT ACATGATGAT GGAATTATCA TTGCTGTCGG AATATTACTT GGTACTTTG 6360
 TTgCACAACG TGCACTAGTT AAAGCAGGAT TACATAAAGA TACTTTAGTA GATATTATTT 6420
 TTTATAGTGC ACTATTTGGA TTTATCGCGG CACGAATCTA TTTGTGATT TTCCAATGGC 6480
 35 CATATTACGC GGAAATCCA AGTGAAATTA TTAAATATG GCATGGTGGA ATAGCAATAC 6540
 ATGGTGGTTT AATAGGTGGC TTTATTGCTG GTGTTATTGT ATGTAAAGTG AAAAATTTAA 6600
 ACCCATTTCA AATTGGTGAT ATCGTTGCGC CAAGTATAAT TTTAGCGCAA GGAATTGGAC 6660
 40 GCTGGGGTAA CTTTATGAAT CACGAGGCAC ATGGTGGATC GGTGTCACGC GCTTTTTTAG 6720
 AACAATTACA TTTGCCTAAT TTTATAATAG AAAATATGTA TATTAACGGC CAATATTATC 6780
 ATCCAACATT CTTATATGAA TCCATTTGGG ATGTCGCTGG ATTTATTATC TTAGTTAATA 6840
 TTCGTAAACA TTTAAATTA GGAGAAACAT TCTTTTTATA TTAACTTGG TATCAATTG 6900
 GTCGATTCTT TATAGAAGGA TTACGTACAG ATAGCTTAAT GCTCACAAGT AATATTAGAG 6960
 50 TTGCACAATT AGTATCAATT CTTTAATTT TAATAAGTAT AAGTTTAATT GTATATAGAA 7020
 GGATTAAGTA TAATCCACCG TTGTATAGCA AAGTTGGGGC GCTTCCATGG CCAACAAAAA 7080

55

	TTATGGCGTG TATACCGTCT TGTTAAATTT TCGAAAGTTT TTAAGAATGT AATTATCATT	7200
	GAATTTTCGA AATTTATTCC AAGTATGGTA CTGAAAAGAC ATATATATAA ACAACTTTTA	7260
5	AATATTAATA TCGGTAATCA ATCGTCGATA GCTTATAAAG TAATGTTAGA TATTTTTTAC	7320
	CCAGAACTGA TTACGATTGG TAGTAACAGT GTTATTGGTT ACAATGTAAC AATTTTGACG	7380
	CATGAAGCAT TAGTTGATGA ATTTCTGTAT GGACCAGTGA CGATAGGATC TAACACTTTG	7440
10	ATTGGTGCAA ATGCTACCAT TTTACCCGGT ATAACGATTG GTGACAATGT AAAAGTTGCA	7500
	GCTGGTACGG TTGTTTCAAA AGATATACCG GATAATGGAT TTGCATATGG CAACCCTATG	7560
	TATATAAAAA TGATTAGGAG GTGACAATTT TATGGCGCAA AAGAATAATA ATGTAATTCC	7620
15	AATGACTTTT GATGATGCAT TTTATCGTAA AATGGCTAAA CAGAAGTTTA AACAAAGAGA	7680
	ATATAAACGA GCTGCTGAAT ACTTTGAAAA AGTGTTAGAA TTGTCACCTG ATGATCTGGA	7740
20	AATTCAAATT GATTATGCAC AATGTCTAGT GCAACTTGGT ATTGCTAAAA AAGCAGAACA	7800
	TTTATTTTAT GACAATATTA TTTATAATAG GCATCTAGAA GATAGCTTTT ATGAATTGAG	7860
	TCAGCTCAAC ATTGAAGTTA ACGAACCAAA CAAGGCATTC TTGTTTGGTA TTAATTATGT	7920
25	TATTGTTAGC GACGACCAAG ATTATAGAGA TGAATTAGAT CAAATGTTTG ATGTGAAATA	7980
	TCAAAGTGAA GAACAAATTG AACTTGAAGC TCAATTGTTT GTAGTTCAA TACTATTCCA	8040
	ATATCTTTTT TCTCAAGGTC GATTAAAAGA TGCAAAGAAT TATGTCTTAC ATCAACCACA	8100
30	AGAAGTTCAA GATCATCGTG TAGTACGTAA TTTATTGGCA ATGTGTTATT TATATCTCGG	8160
	TGAATATGAT ACgGCTAAAG CATTGTACGA aGCACTATTA CAAGAGGATA GTACaGATAT	8220
	ATATGCATTA TGCCATTATA CTTTGCTACT TTATAACACT AAGGAAAATG AACAAATATCA	8280
35	AAAATATTTA AAAATATTAA ACAAAGTTGT ACCTATGAAT GACGATGAAA GTTTTAAATT	8340
	AGGTATTGTA TTAAGTTATT TAAAGCAGTA TCGTGCAATCA CAACAATTGT TGTACCCTTF	8400
40	ATATAAAAAA GGGAAATTTT TATCAATTCA AATGTACAAT GCTTTAGCAT ATAATTATTA	8460
	TTATTTAGGT GAAGAAGACG AAAGTCATTA CTACTGGGAT AAATTGAAGC AAATTTCTAA	8520
	AGTGGAATTT GGACATGCGC CTTGGGTAAT TGAAAATAGC AAAGAAGTTT TTGACCAACA	8580
45	TATTTTGCCA TTACTTCAAA GTGATGACAG TCATTATCGT TTATATGGTA TTTTTTTATT	8640
	GGATCAATTA AATGGTAAAG AAATGTGAT GACGGAAAGT ATTTGGCAGG TTTTGGAAAA	8700
	TCTAAATAAT TATGAGAAAT TGTATTTAAC GTATTTAGTT CAAGGTTTAA CGCTCAATAA	8760
50	ATTAGACTTC ATTCATCGCG GCTTATTAAC GCTTTACCAT AATGAATTAT TTGTAAGTGA	8820
	AAATGATGTA ATGTTTGCAT GGATTAATCA AGGTGAACTC ATAATTGCTG AAAAAGTAGA	8880

55

	TCGAAACGTT	ACAAAGAAGC	AAATTACAAC	ATGGTTAGGC	ATAACACAAT	ATAAACTGAA	9000
	CAAAATGATT	GAATTTCTCT	TGAGCATATA	GATTTATGAA	AAGTTAGATT	TATTATATAA	9060
5	TGCGCATAAT	GATTAATAAT	GAGGAGGCGT	TAATAAAATG	ACTGAAATAG	ATTTTGATAT	9120
	AGCAATTATC	GGTGCAGGTC	CAGCTGGTAT	GACTGCTGCA	GTATACGCAT	CACGTGCTAA	9180
	TTTAAAAACA	GTTATGATTG	AAAGAGGTAT	TCCAGGCGGT	CAAATGGCTA	ATACAGAAGA	9240
10	AGTAGAGAAC	TTCCCTGGTT	TCGAAATGAT	TACAGGTCCA	GATTTATCTA	CAAAAATGTT	9300
	TGAACACGCT	AAAAAGTTTG	GTGCAGTTTA	TCAATATGGA	GATATTAAAT	CTGTAGAAGA	9360
	TAAAGGCGAA	TATAAGTGA	TTAACTTTGG	TAATAAAGAA	TTAACAGCGA	AAGCGGTTAT	9420
15	TATTGCTACA	GGTGCAGAAT	ACAAGAAAAT	TGGTGTTCCT	GGTGAACAAG	AACCTTGGTGG	9480
	ACGCGGTGTA	AGTTATTGTG	CAGTATGTGA	TGGTGCAATC	TTTAAAAATA	AACGCCTATT	9540
20	CGTTATCGGT	GGTGGTGATT	CAGCAGTAGA	AGAGGGAACA	TTCTTAACTA	AATTTGCTGA	9600
	CAAAGTAACA	ATCGTTCACC	GTCGTGATGA	GTTACGTGCA	CAGCGTATTT	TACAAGATAG	9660
	AGCATTCAAA	AATGATAAAA	TCGACTTTAT	TTGGAGTCAT	ACTTTGAAAT	CAATTAATGA	9720
25	AAAAGACGGC	AAAGTGGGTT	CTGTGACATT	AACGTCTACA	AAAGATGGTT	CAGAAGAAAC	9780
	ACACGAGGCT	GATGGTGTAT	TCATCTATAT	TGGTATGAAA	CCATTACAG	CGCCATTTAA	9840
	AGACTTAGGT	ATTACAAATG	ATGTTGGTTA	TATTGTAACA	AAAGATGATA	TGACAACATC	9900
30	AGTACCAGGT	ATTTTTCAG	CAGGAGATGT	TCGCGACAAA	GGTTTACGCC	AAATTGTCAC	9960
	TGCTACTGGC	GATGGTAGTA	TTGCAGCGCA	AAGTGCAGCG	GAATATATTG	AACATTTAAA	10020
	CGATCAAGCT	TAATTCGAAG	TCGAATTAAG	ATGTTGAGCT	GTAAATTATT	TGGATATTTA	10080
35	TTTTAATAGT	GTCATCACAG	CGTTAAAATA	ATGCTTACT	TTTAAATTAA	AGCAAATTAT	10140
	ATAGTAAACT	AGAACTTAGT	ACGTATCATT	TGTGCGTTTC	AATGAGTTCT	AGTTTTTTTA	10200
40	TATGTTATAT	TAAACTTATA	ACTTTATGGG	AGTGGGACAG	AAATGATAAA	GAGCCACTAA	10260
	TGATTTATTA	TGTAGTGGTT	CTTAAACATT	AGCCACAGCT	AATGTGTACT	TAAAAATAGG	10320
	AATACATGAG	TAAAACTCAT	GCATAAGAAA	TACTAATTTC	TATAGAAAAA	GTATTACTTT	10380
45	ATCGTTGTCC	CACCCCAACT	TGCACATTAT	TGTAAGCTGA	CTTTCCGCCA	GCTTCTGTGT	10440
	TGGGGCCCCG	CCAACCTGCA	CATTATTGTA	AGCTGACTTT	TCGTCAgCTT	CTGTGTTGGG	10500
	GCCCCGCCAA	CTTGACATT	ATTGTAAGCT	GACTTTTCGT	CAGCTTCTGT	GTTGGGGCCC	10560
50	CGCCAACTTG	CATTGTCTGT	AGAAATTGGG	AATCCAATTT	CTCTATGTTG	GGGCCCACAC	10620
	CCCAACTCGC	ATTGCCTGTA	GAATTTCTTT	TCGAAATTCT	CTGTGTTGGG	GCCCACACCC	10680

55

	ACTCGCATTG CCTGTAGAAT TTCTTTTCGA AATTCTCTGT GTTGGGGCCC CTGACTAGAG	10800
	TTGAAAAAAG CTTGTTGCAA GCGCATTTTC ATTCAGTCAA CTACTAGCAA TATAATATTA	10860
5	TAGACCCTAG GACATTGATT TATGTCCCAA GCTCCTTTTA AATGATGTAT ATTTTITAGAA	10920
	ATTTAATCTA GACATAGTTG GAAATAAATA TAAACATCG TTGCTTAATT TTGTCATAGA	10980
	ACATTTAAAT TAACATCATG AAATTCGTTT TGGCGGTGAA AAAATAATGG ATAATAATGA	11040
10	AAAAGAAAAA AGTAAAAGTG AACTATTAGT TGTAACAGGT TTATCTGGCG CAGGTAAATC	11100
	TTTGTTTATT CAATGTTTAG AAGACATGGG ATATTTTGT GTAGATAATC TACCACCACT	11160
	GTTATTGCCT AAATTTGTAG AGTTGATGGA ACAAGGAAAT CCATCCTTAA GAAAAGTGGC	11220
15	AATTGCAATT GATTTAAGAG GTAAGGAACT ATTTAATTCA TTAGTTGCAG TAGTGGATAA	11280
	AGTCAAAAAGT GAAAGTGACG TCATCATTGA TGTATGTTT TTAGAAGCAA GTACTGAAAA	11340
20	ATTAATTTCA AGATATAAGG AAACGCGTCG TGCACATCCT TTGATGGAAC AAGGTAAAAG	11400
	ATCGTTAATC AATGCAATTA ATGATGAGCG AGAGCATTTG TCTCAAATTA GAAGTATAGC	11460
	TAATTTTGTT ATAGATACTA CAAAGTTATC ACCTAAAGAA TTAAGAAGAAC GCATTCGTCG	11520
25	ATACTATGAA GATGAAGAGT TTGAACTTT TACAATTAAT GTCACAAGTT TCGGTTTTAA	11580
	ACATGGGATT CAGATGGATG CAGATTTAGT ATTTGATGTA CGATTTTAC CAAATCCATA	11640
	TTATGTAGTA GATTTAAGAC CTTTAACAGG ATTAGATAAA GACGTTTATA ATTATGTTAT	11700
30	GAAATGGAAA GAGACGGAGA TTTCTTTGA AAAATTAACT GATTGTAG ATTATATGAT	11760
	ACCCGGGTAT AAAAAAGAAG GGAAATCTCA ATTAGTAATT GCCATCGGTT GTACGGGTGG	11820
	ACAACATCGA TCTGTAGCAT TAGCAGAACG ACTAGGTAAT TATCTAAATG AAGTATTTGA	11880
35	ATATAATGTT TATGTGCATC ATAGGGACGC ACATATTGAA AGTGGCGAGA AAAAATGAGA	11940
	CAAAATAAAG TTGTACTTAT CGGTGGTGGC ACTGGCTTAT CAGTTATGGC TAGGGGATTA	12000
	AGAGAATTCC CAATTGATAT TACGGCGATT GTAACAGTTG CTGATAATGG TGGGAGTACA	12060
40	GGGAAAATCa GAGATGAAAT GGATATACCA GCACCAGGAG ACATCAGAAA TGTGATTGCA	12120
	GCTTTAAGTG ATTCTGAGTC AGTTTAAAGC CAACTTTTTT AGTATCGCTT TGAAGAAAAT	12180
45	CAAATTAGCG GTCATCATT AGGTAATTTA TTAATCGCAG GTATGACTAA TATTACGAAT	12240
	GATTTCCGAC ATGCCATTAA AGCATTAAAGT AAAATTTTAA ATATTAAAGG TAGAGTCATT	12300
	CCATCTACAA ATACAAGTGT GCAATTAAAT GCTGTTATGG AAGATGGAGA AATTGTTTTT	12360
50	GGAGAAACAA ATATTCCTAA AAAACATAAA AAAATTGATC GTGTGTTTTT AGAACCTAAC	12420
	GATGTGCAAC CAATGGAAGA AGCAATCGAT GCTTTAAGGG AAGCAGATTT AATCGTTCTT	12480

55

GCGTTAATTC ATTCTGATGC GCCTAAGCTA TATGTTTCTA ATGTGATGAC GCAACCTGGG 12600
 GAAACAGATG GTTATAGCGT GAAAGATyAT ATCGATGCGA TTCATAGACA AGCTGGACAA 12660
 5 CCGTTTATTG ATTATGTCAT TTGTAGTACA CAAACTTTCA ATGCTCAAGT TTTGAAAAAA 12720
 TATGAAGAAA AACATTCTAA ACCAGTTGAA GTTAATAAGG CTGAACTTGA AAAAGAAAGC 12780
 ATAAATGTAA AAACATCTTC AAATTTAGTT GAAATTTCTG AAAATCATTT AGTAAGACAT 12840
 10 AATACTAAAG TGTTATCGAC AATGATTTAT GACATAGCTT TAGAATTAAT TAGTACTATT 12900
 CCTTTCGTAC CAAGTGATAA ACGTnAATAA TATAGAACGT AATCATATTA TGATATGATA 12960
 ATAGAGCTGT GAAAAAATG AAnATAGACA GTGGTTCTAA GGTGAATCAT GTTTTAAATA 13020
 15 AGAAAGGAAT GACTGTACGA TGAGCTTTC ATCAGAAATG AAAATGAAT TAACTAGAAT 13080
 AGACGT 13086

20 (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1350 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

30 CATTAGTCAT GAAATAGCC GACAACTTCA TCTGTGAAAT CACCGGCCTT TTATTTTAGC 60
 TAACTTTATT TCTGATTTTA CGATTTTAAT TGATCATACA GAGAAAGTGA TCTTTTACA 120
 ATTTCTAAAA ACTCATGATC TATATTGGAC ATTTGATGAA AATAAGACAA AATGTTTCT 180
 35 GTTAGCTTCT CTGTGTTTGG GAATGAATCA TCTTCTTTAA TCCAAATCGC TAATTCGCCT 240
 AATGGTGTGTT TATCATCTTT AAATGTTTGT ATATATTCGT AAAAGCTCAT AGTATTCCTT 300
 CTCTCAATTT ACTTATATAA ATCCTACCAC GAAAGCTTTC AAGAAAACAC AATTAAATGT 360
 40 CTATTTAGTG AACTTTTAA GGTGTGCAC TCTTTTAATG TCTGCCAATT AGGTCAATTA 420
 ATCATCACAA TGTACAATTA ACTCTATTTT CAGTTCATAT ACTCACACAC CGTTTTGAA 480
 CAACACATTA ACTTCTCATT TAGATAAAAC GCAAAAAGC CTGGCACCAA TACAATAGAT 540
 GCCAGACTAA GAGTCTACTA TATAAATTTA TTTAGCGTAT GGTTTTACTT CGATTGCACC 600
 TTCATTTTCA TCATGAACAC CATGCTTATA ATAATCAATA TATTGTGGCT CTAAAGGCTT 660
 50 TCTGCCACGT ATAATGTCTG CTGCTTTTTT AGCTAACATT AAAACAGGTG CGTGTATATT 720
 GCCATTTGTC GTACGTGGCA TAGCTGATGC ATCAACTACA CGTAAATTTT CCATACCGTG 780

55

ACTACAAGAT GGGTGAATG CTGTTTCACC ATCTCTACGA ACCCAATCAA GAATTTCTTC 900
 GTCTGTTTGC ACTTCTGGTC CTGGTGAAAT TTCTCCACCA TTGAATGGAT CCATTGCTTT 960
 5 TTGAGATAAG ATATTTCTTG CTACACGAAT TGCTTCTACC CATTCTTTTT TATCTTCTTC 1020
 TGTTGATAAA TAATTAAAGC GGATACTTGG TTTTTCGAAT GGATCTTTAG ATTTGATTTT 1080
 CAAGCTACCA CGAGAGTTTG AATACATTGG TCCTACGTGA ACTTGATAAC CATGTGCGAC 1140
 10 CGCTGCCTTT TGACCATCAT ATCTTACAGC TATTGGTAAG AAATGGAACA TTAAGTTAGG 1200
 ATAAcCAACT TCGTTATTTG AACGTACAAA TCCGCCACCT TCAAAATGGT TAGATGCTGC 1260
 15 TGCACCTGTA CGTGTGAAAA TCCATTGTAA ACCAATAAAT GGcATGCGCT TGAtATCTAA 1320
 GCTTGGctGt AATGATACAG GTTCCTTACA 1350

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1376 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TAATGCTATT GGCAACACCA TATATGAAAn CTCCAAACGA TCCTAAACCG ACTATAGATT 60
 30 CACCAAATTT nACAATCCAT GAATAAAGTA GTGGCCATAA GAATAACAAT ATGACAACTA 120
 AAAATGTACA GTAAATGCA GTCATAATTG GAACTAGACG TTTACCACTA AAAAATGATA 180
 ATGCTAATGG TAATTCTGTT TCACTAAACT TATTGTATGC ATAAGCTGCT ATTAAACCTA 240
 35 TTACAATACC AACAAAGACA TTGCCATTAT TCATCTTTTC AAAAGCTGAA TTTATTTCCG 300
 ArGCTTTTCAT TCCTAATAAA GGCGCTAATT TCATTGGTGA TAATACAACT GTAACATAAA 360
 40 AATATCCTAA CGTrGCTGCA rGCGsGACTG CACCATCAIT TTTCTTTGCC ATTCCTATAG 420
 CTACACCAAT TGCAAATAAA ATACCTAATT GCTCTAAAAT CGTAGTACCT ACCGTAGTAA 480
 AGAACATTGC GATTTTCGGC GTCGCATGAA GTGCATTTAA CGTATTACCA ATTCCGGCAA 540
 45 TAATTGCTGC AGCCGGTAAA ATGGCAACTG GTAACATTAA CGAACGCCCT AAATTTTGGA 600
 AAAATTTATA CATTGAATGT CATCCTTCTT AAAATAATGT AGAAATATAA AGATTACTAA 660
 TGTAAC TAGA ATAAC TACTT CGATACTCCG TTATAGTCAC CTAGGCTTAC TAACCAGCTA 720
 50 TATTTCTACC TCAAGTTATT TTATAAACTT TTTACAATTT CATGCAATTC TTGTTGTAAC 780
 TTTGCTGTTT GTGTTTCAAT CTCTTTTGTA ATATAATCGA TACGCTCGTT TCGTTTTAAA 840

AAAGACCGTG AATCTTAGTA GGACCAACAT AAGCAACAGG TAATATTGGT GACTTACTTA 960
 ACATTGCAAT TGTGAAGCA CCaCGTTTCA AAGGTGCACC TTCTTGCGAT GTGCGAGAAC 1020
 5 CTGTTGGGAA GATACCAACT GTCTTATTAT CTTTCAACAA ATTGATTGGG CGTTTTAAAG 1080
 TACTAGGTCC TGGATTTTCA CGATCTACAG GAAATGCATT TAAAGACGTT AAAAAATTAC 1140
 CAATCCATTT ATTTTGAAT AATTCTTTT TAGCCATATA ATGAATTTGA TTAGGATATA 1200
 10 ATGCCATACC TAGCATAATG ACTTCGTTAT AACTTTTCATG CGTACAAGTT ACGACATATT 1260
 TACTATCCTT AGGAATATTA TCTTTACCGA TTACGTATAA TGATTTTGAC ATTTTAACTA 1320
 AAATGAAATT CAAAATCTTA CTAATCACTG AATACATTGT GCCACCTACT TAACTT 1376
 15

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7363 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TTGTCATACC AATATTTTGT AAAATATGGA ACACAAGTAA AGTGACGAAA CCAACGATAA 60
 AGATTTTGTT AAATTGATCT TCAATTTTCG CAGCTAATCT TATTAGATGG AAGATTAAAA 120
 30 ATAAAAATAT TAAGATCAAT ATGACAGAAC CGATAAAGCC AAGTTCCTCT CCAATCACTG 180
 AAAAGATAAA GTCAGTATGA TTTTCAGGTA TATAAACTTC ACCGTGATTG TATCCTTTAC 240
 CTAGTAACTG TCCAGAACCG ATAGCTTTAA GTGATTCAGT TAAATGaTAG CCATCACCAC 300
 35 TACTATATGT ATAGGGGTCA AGCCATGAAT TGATTCGTCC CATTGATAC AGTTGGaCAC 360
 CTAATAAATT TTCAATTAAT GCGGGTGCAT ATAGaATACC TAAATGACT GTCATTGCAC 420
 CAACaATACC TGTAATAAAG ATAGGTGCTA AGATACGCCA TGTATACCA CTTACTAACA 480
 40 TCACACCTGC AATAATAGCA GCTAATACTA ATGTAGTTCC TAGGTCATTT TGCAGTAATA 540
 TTAAAATACT TGGTACTAAC GAGACACCAA TAATTTTGAA AAATAATAAC AAATCACTTT 600
 GGAATGATTT ATTGAATGTG AATTGATTAT GTCTAGAAAC GACACGCGCT AATGCTAAAA 660
 45 TTAAAATAAT TTTCATGAAT TCAGATGGCT GAATACTGAT AGGGCCAAAC GTGTACCAAC 720
 TTTTGGCACC ATTGATAATA GGTGTAATAG GTGACTCAGG AATAACGAGC AAGCCTATTA 780
 50 ATAATAGACA GATTAAGAAA TACAATAAAT ATGTATAATG TTTAATCTTT TTAGGTGAAA 840
 TAAACATGAT GATACCTGCA AAAATTGCAC CTAAAATGTA ATAAAAAATT TGTCTGATAC 900

	TTGCTAAAAC AGCTATAGTG GCTACTAATA CCCAGTCTAC TTTGCGAAnC aATGCTTATC	1020
	CGGCTGTTGA CGAGATGAAT AATTCATTGC AAACTCCTTT TATACTCACT AATGTTTATA	1080
5	TCAATTTTAC ATGACTTTTT AAAAATTAGC TAGAATATCA CAGTGATATC AGCTATAGAT	1140
	TTCAATTTGA ATTAGGAATA AAATAGAAGG GAATATTGTT CTGATTATAA ATGAATCAAC	1200
10	ATAGATACAG ACACATAAGT CCTCGTTTTT AAAATGCAAA ATAGCATTAA AATGTGATAC	1260
	TATTAAGATT CAAAGATGCG AATAAATCAA TTAACAATAG GACyAAATCA ATATTAATTT	1320
	ATATTAAGGT AGCAAACCCT GATATATCAT TGGAGGAAAA CGAAATGACA AAAGAAAATA	1380
15	TTTGTATCGT TTTTGGAGGG AAAAGTGCAG AACACGAAGT ATCGATTCTG ACAGCACAAA	1440
	ATGTATTAAT TGCAATAGAT AAAGACAAAT ATCATGTTGA TATCATTTAT ATTACCAATG	1500
	ATGGTGATTG GAGAAAGCAA AATAATATTA CAGCTGAAAT TAAATCTACT GATGAGCTTC	1560
20	ATTTAGAAAA TGGAGAGGCG CTTGAGATTT CACAGCTATT GAAAGAAAGT AGTTCAGGAC	1620
	AACCATACGA TGCAGTATTC CCATTATTAC ATGGTCCTAA TGGTGAAGAT GGCACGATTC	1680
	AAGGGCTTTT TGAAGTTTTG GATGTACCAT ATGTAGGAAA TGGTGTATTG TCAGCTGCAA	1740
25	GTTCTATGGA CAAACTTGTA ATGAAACAAT TATTTGAACA TCGAGGGTTA CCACAGTTAC	1800
	CTTATATTAG TTTCTTACGT TCTGAATATG AAAAATATGA ACATAACATT TTAAATTAG	1860
	TAAATGATAA ATTAAATTAC CCAGTCTTTG TTAAACCTGC TAACTTAGGG TCAAGTGTAG	1920
30	GTATCAGTAA ATGTAATAAT GAAGCGGAAC TTAAAGAAGG TATTAAAGAA GCATTCCAAT	1980
	TTGACCGTAA GCTTGTTATA GAACAAGGCG TTAACGCACG TGAAATTGAA GTAGCAGTTT	2040
	TAGGAAATGA CTATCCTGAA GCGACATGGC CAGGTGAAGT CGTAAAAGAT GTCGCGTTTT	2100
35	ACGATTACAA ATCAAAATAT AAAGATGGTA AGGTTCAATT ACAAATTCCA GCTGACTTAG	2160
	ACGAAGATGT TCAATTAACG CTTAGAAATA TGGCATTAGA GGCATTCAAA GCGACAGATT	2220
40	GTTCTGGTTT AGTCCGTGCT GATTTCTTTG TAACAGAAGA CAACCAAATA TATATTAATG	2280
	AAACAAATGC AATGCCTGGA TTTACGGCTT TCAGTATGTA TCCAAAGTTA TGGGAAAATA	2340
	TGGGCTTATC TTATCCAGAA TTGATTACAA AACTTATCGA GCTTGCTAAA GAACGTCACC	2400
45	AGGATAAACA GAAAAATAAA TACAAAATTG ACTAACTGAG GTTGTTATTA TGATTAATGT	2460
	TACATTAAAG CAAATTCAAT CATGGATTCC TTGTGAAATT GAAGATCAAT TTTTAAATCA	2520
	AGAGATAAAT GGAGTCACAA TTGATTCACG AGCAATTTCT AAAAATATGT TATTTATACC	2580
50	ATTTAAAGGT GAAAATGTTG ACGGTCATCG CTTTGTCTCT AAAGCATTAC AAGATGGTGC	2640
	TGGGGCTGCT TTTTATCAAA GAGGGACACC TATAGATGAA AATGTAAGCG GGCCTATTAT	2700

55

	AAACCCTAAA GTAATTGCCG TCACAGGGTC TAATGGTAAA ACAACGACTA AAGATATGAT	2820
	TGAAAGTGTA TTGCATACCG AATTTAAAGT TAAGAAAACG CAAGGTAATT ACAATAATGA	2880
5	AATTGGTTTA CCTTTAACTA TTTTGAATT AGATAATGAT ACTGAAATAT CAATATTGGA	2940
	GATGGGGATG TCAGGTTTCC ATGAAATTGA ATTTCTGTCA AACCTCGCTC AACCAGATAT	3000
10	TGCAGTTATA ACTAATATTG GTGAGTCACA TATGCAAGAT TTAGGTTTCGC GCGAGGGGAT	3060
	TGCTAAAGCT AAATCTGAAA TTACAATAGG TCTAAAAGAT AATGGTACGT TTATATATGA	3120
	TGGCGATGAA CCATTATTGA AACCACATGT TAAAGAAGTT GAAAATGCAA AATGTATTAG	3180
15	TATTGGTGTT GCTACTGATA ATGCATTAGT TTGTTCTGTT GATGATAGAG ATACTACAGG	3240
	TATTTCAATTT ACGATTAATA ATAAAGAACA TTACGATCTG CCAATATTAG GAAAGCATAA	3300
	TATGAAAAAT GCGACGATTG CCATTGCGGT TGGTCATGAA TTAGGTTTGA CATATAACAC	3360
20	AATCTATCAA AATTTAAAAA ATGTCAGCTT AACTGGTATG CGTATGGAAC AACATACATT	3420
	AGAAAAATGAT ATTACTGTGA TAAATGATGC CTATAATGCA AGTCCTACAA GTATGAGAGC	3480
	AGCTATTGAT ACACTGAGTA CTTTGACAGG GCGTCGCATT CTAATTTTAG GAGATGTTTT	3540
25	AGAATTAGGT GAAAATAGCA AAGAAATGCA TATCGGTGTA GGTAATTATT TAGAAGAAAA	3600
	GCATATAGAT GTGTTGTATA CGTTTGGTAA TGAAGCGAAG TATATTTATG ATTCGGGCCA	3660
	GCAACATGTC GAAAAAGCAC AACACTTCAA TTCTAAAGAC GATATGATAG AAGTTTTAAT	3720
30	AAACGATTTA AAAGCGCATG ACCGTGTATT AGTTAAAGGA TCACGTGGTA TGAAATTAGA	3780
	AGAAGTGGTA AATGCTTTAA TTTCATAGAG ATTAGTCGAG GGACCTTTTA CTTATAAAAA	3840
	TGATTTGAAT TAATACTAAA AGATTACAAA GAAGAGGTGG TTTTGTGTGT AAATACAAAA	3900
35	TTGCCTTTTT CTTTTTATGT TAAATCTATA AATTTGAAAC TAAATCAAGG TTAATTCTAT	3960
	GTACACACTT TATATAGGAA GTAGTTTGAA TGTTTATATA ATGTTTTACA AAAAGATGTA	4020
40	GTATTATAAT GTCTAATTTT ACATGTGTTT CAGTAAATT TGTGTGGAA TGTTAACGAT	4080
	ATACGTATTT TATAAAAAAT TTTTATAAT GATTATTCGA ATGATGCGTA ACGCTTACAT	4140
	CTTATCTAAT GCTAGCTTTT TGACAAAAAT ATGACAATCA ATTAATGTGA TTCTAATAAA	4200
45	TATTCGCAAA TTGCTTTATT GCGATTAAAT TTTTTTGGTG GTACTATATA GAAGTTGATG	4260
	AAATATTAAT GAACTTATAT GCAAAAGTAT ATTGAGAAAT AACAGGTAA AAAGGAGAAT	4320
	TATTTTGCAA AATTTTAAAG AACTAGGGAT TTCGGATAAT ACGGTTCACT CACTTGAATC	4380
50	AATGGGATTT AAAGAGCCGA CACCTATCCA AAAAGACAGT ATCCCTTATG CGTTACAAGG	4440
	AATTGATATC CTTGGGCAAG CTCAAACCGG TACAGGTAAA ACAGGAGCAT TCGGTATTCC	4500

55

	AGAATTGGCA ATGCAGGTAG CTGAACAATT AAGAGAATTT AGCCGTGGAC AAGGTGTCCA	4620
	AGTTGTTACT GTATTCGGTG GTATGCCTAT CGAACGCCAA ATTAAAGCCT TGAAAAAAGG	4680
5	CCCACAAATC GTAGTCGGAA CACCTGGGCG TGTTATCGAC CATTTAAATC GTCGCACATT	4740
	AAAAACGGAC GGAATTCATA CTTTGATTTT AGATGAAGCT GATGAAATGA TGAATATGGG	4800
10	ATTCATCGAT GATATGAGAT TTATTATGGA TAAAATTCCA GCAGTACAAC GTCAAACAAT	4860
	GTTGTTCTCA GCTACAATGC CTAAAGCAAT CCAAGCTTTA GTACAACAAT TTATGAAATC	4920
	ACCAAAAATC ATTAAGACAA TGAATAATGA AATGTCTGAT CCACAAATCG AAGAATTCTA	4980
15	TACAATTGTT AAAGAATTAG AGAAATTGTA TACATTTACA AATTCCTAG ATGTTTCATCA	5040
	ACCTGAATTA GCAATCGTAT TCGACGTAC AAAACGTCGT GTTGATGAAT TAACAAGTGC	5100
	TTTGATTTCT AAAGGATATA AAGCTGAAGG TTTACATGGT GATATTACAC AAGCGAAACg	5160
20	TtAGAAAGTA TtanAGAAAT TTAAAAATGA CCAAATTAAT ATTTTAGTCG CTAATGATGT	5220
	AGCAGCaAGA GGACTAGATA TTTCTGGTGT GAGTCATGTT TATAACTTTG ATATACCTCA	5280
	AGATACTGAA AGCTATACAC ACCGTATTGG TCGTACGGGT CGTGCTGGTA AAGAAGGTAT	5340
25	CGCTGTAACG TTGTGTAATC CAATCGAAAT GGATTATATC AGACAAATTG AAGATGCAAA	5400
	CGGTAGAAAA ATGAGTGCAy TcGTCCACCA CATCGTAAAG AAGTACTTCA AGCACGTGAA	5460
	GATGACATCA AAGAAAAAGT TGAAAACCTGG ATGTCTAAAG AGTCAGAATC ACGCTTGAAA	5520
30	CGCATTCTTA CAGAGTTGTT AAATGAATAT AACGATGTTG ATTTAGTTGC TGCACTTTTA	5580
	CAAGAGTTAG TAGAAGCAAA CGATGAAGTT GAAGTTCAAT TAACTTTTGA AAAACCATTA	5640
	TCTCGCAAAG GCCGTAACGG TAAACCAAGT GGTTCCTCGTA ACAGAAATAG TAAGCGTGGT	5700
35	AATCCTAAAT TTGACAGTAA GAGTAAACGT TCAAAAGGAT ACTCAAGTAA GAAGAAAAGT	5760
	ACAAAAAAAT TCGACCGTAA AGAGAAGAGC AGCGGTGGAA GCAGACCTAT GAAAGGTCGC	5820
40	ACATTTGCTG ACCATCAAAA ATAATTTATA GATTAAGAGC TTAAAGATGT AATGTCTTGA	5880
	GCTCTTTTTT GTTTTCAATA ATTGATTCTC TGTAAGATATC aAAGTaCTAA CGTTTTAAAG	5940
	GTTAAATATT TAATTGGATT GAGATCTGTA TGCGGTTATA TCaTTCTGTG TAAATATGGT	6000
45	TCTCCACCAA ATGTGGTGAG TATATAATTT AAAGAACTAT TTTTAAATTA AGAATAATCG	6060
	AACATAAATA AACTTTATGA AATTTAGTA TCATGTTCTT ATAAAAACA ATAGGGCTTT	6120
	TTGctGACGC TAGTGCGCGA TAAATAATAA GTTGAATATA AAAAAGATCA CTGCCAATCA	6180
50	TTCGTTTAAT GGCAGCGATC TTTTTTATTT AATTATTTCT CTTTCCACTG CAACATTTGA	6240
	TAACCAATGC GTGGATGTGT TTTAATAATA TCTTTTTCGT CCTCATGACA TTGTGAAAGT	6300

55

CCATATATTC GTTTTAATAT CATCTCATAA GTGAGTACTT TTCCTTTATG ATTTGACAAT 6420
 AGTTCTAACA AGCTAAATTC ATTTGGCGTC AAATGTACCT CCGATTATT AATAACAACA 6480
 5 GATTTGGAGC CAAAGTCGAT GCTTAGCAAA CCGTTAGTAA ATACAATGTT AGTTTCTTGA 6540
 TGTGACTTAG CGATTCTCTC GATGACTCGT ATTCGTGCCC GAAGCTCATC AACATTAAAA 6600
 GGTTTAGTCA TATAGTCATT CGCACCGTTA TCTAAAGCTT GAATAATTGT TTGTTCTTCT 6660
 10 TGTCTTGCAC TTATTACAAT GATAGGAATG TCAGTATGTT GCCTGATTTC TGAAATCAAA 6720
 CATAATCCAT CTTTATCTGG TAAACCTAAA TCTAATAAAA TGACATCTGG TTTATCAATT 6780
 TGAATTTTAA AGTGTGCTTG TGTGGCATTG TCGGCTGTAG TTACATTGTA ATAATCTAAA 6840
 15 GTTAATGCAA CATCAAGTAA ATGTGTGATT GCGTGATCAT CTTCAATTAT CAATATTTTA 6900
 GATTGCATTA TACGTCTCCT TCGTTAAAGT CTGTATATAT ATTGAAATAG AATATACTGC 6960
 CGTGTGGTTG GTTCGGTTTA TATTGTAAGT TTGATTGATG TTTGTGTAGG ATAGTCTGTA 7020
 CTAATATATA GCCTAGTCCC ATGCTTTCTT TTTGGTTATC TTTAAAATAT TTATTTGATC 7080
 CTGTGTAAAA AGGCTCGAAT ATCTTTTGTT GTTCTTCTAA ACTAATTCCA GGTCCTTCGT 7140
 25 CTATAACGGC AAATTCGATT TGTCATAGC TAGCATAACG AATAGATAAA TTGATTTTGG 7200
 TGTCAAGTAG AGTGTGTTTA ACTGCATTTT CAATCAAATT GAATAAAGCT TGTAATAATCA 7260
 ACTTACTGTC AATGTGTATA AACTGTAAAT TTAAGTGAAG TGATACAGTT ATACGCTTTT 7320
 30 TTAAATGGCG ACGTTCTAAA ATACATATCG ATTTCTTATA CTA 7363

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10470 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAACAATCG ATAACCACAA TACTTCTATT GTAATTGTTT AACGATTTCTn CGATTAAAAAT 60
 45 CATCTAAATC GTCTGGTACT CGACTTGTTA CAATATTGTT GTCTACAcTa CTGACTCATC 120
 AACTACATGT GCGCCTGCAT TTGATAAATC TTTGCGTACA TTTAATACTG CTGTTAACGT 180
 ACGACCTTTT AAATCGTCTG TATCTATTAG TATTTGTGGC CCATGACAAA TGGCAAATGT 240
 50 TGGTACATCA TTTTITAGTAA AGTATTTAGC AAATGTGCCA TATCGACCTT CTGTATCTCC 300
 ACGTAAATGA TCTGGTGAAA ATCCTCCAGG AATTAATAAT GCATCATAAT CTTCTGGTTT 360

	ATTTGCAGTA TCTCCAATCA CTACAGTATT AAAGCCTGCA TTCTCTAATG CCTCTTTAGG	480
	GCTTGAATAT TCTATATCTT CAAATTCGTT TGCTAGAATA ATTGCTACTT TTTTAGTCAT	540
5	TGAAAATCAC CTTTCTATAT ATCATTGATA TAATTACTAT AGACAAGTAA ATCAGTGATT	600
	AAACATACAA GATATAAAAA ATATTAAGCG ACTGTCGCGA TATCTAACCC TAACACATCT	660
10	TATGTGGCAT TTACTIONAGAT ACTAATTTAA CCTTTTCTTC AAGCTGATCT AACAAATCCAA	720
	TCCATTGATC TATATCTTCA ACACGTACTT CATCAGGATT TACATGATCG ATATCCTCAA	780
	TAAACTTATT TAAACGCGCT TTTATCTGTT CGATTGTTTG CTGTTCAATC ATAAAAAGTT	840
15	AACTCCTTTT ATTTTGTTTT CTTTTTCATT ATTATCCTAA CAGAAATTGC GTTAAAGCGA	900
	TATAATCTTA GCTATATTTA TGACATTCAA ATTATTTTGA CTTTTAAAAA TCCCTTTTTC	960
	AATTAACATA AATTAAGAGA TAATTTGTTA CGAGTGATAA TACGAAGkGG TaTCATACCG	1020
20	ATATGAACCA AATAGAAAGA AGGAAGTTTA AGACGATGAA TAGCGTCAAA TTGAAGCAAC	1080
	CTGTTAGCAT TTACAATGAT CCATGGGAAG TGAAATTTAT ATACATTTAA ATTTTCATGAG	1140
	ACAATAAACG TTGATTTAAT GCGTTTTTTT GCCTTTTTTA TTTTCCTTAT TTTTCTGTT	1200
25	TTACAACAAA ATGGTATCAA AAATGGTATC ATTTGTAGTT ATTTTAGCTT CACATATTAA	1260
	AACAACCACA CTCCTAAATT AATAGGTGGT GTGGTTTTGT TGGTTGTGTG GGGATAAAAA	1320
	TAACCGCATC AGTTAAGATG CGGTTATCTA GCAAGGGCCA CGTATTTATA AATACGTTTA	1380
30	GAATCTCTTC GGCAACTTTG CTATAGACAG TCTATGCTGT TACTAAATTA TACCACCACA	1440
	CAAACCTACT CCCATTGAGG AACACAGAGC TTTGTCGCTC GTCAGCAACG TCATATGAAT	1500
35	TCTCAGTTCA TGTTGTGGTG AACTTTTAAA CGGTCTGTGC CAGTAGCGAC CGAGTCATTT	1560
	CAAGAATGAC CATTTACAT TTATATTATA AACTTTGTCTG TCGTAACTG TATAGTTTTT	1620
	CAGTTGTATT TAAAGTTAAG TTATCTACTT CGCGCTTTCC TTGCCTTAAT TGTGAAATTA	1680
40	CATATTGCGC TACGCCAGTT TGTTTGTGAA TTTGGTAACC TGTTATATCA CTTTGTATCA	1740
	ATTCAATTAT TTTTAATTTA TAATCACTCA TATTATCTAC GTCCATTCTT TTTATCTAAA	1800
	CAATAAAAAAT GTGTCTTTCT CCCGATAAAT AATAACAATG GTAGGCTTAA TAAAAACAAT	1860
45	ATTAAATACA TTTGTTCTGT CATAATTGAA AACCTCCAAA TAATATTATA TTATATAAGT	1920
	GTAAGGAGGA GCCATCAGGC TCCAAGCATA ATGTTAATCT TTGTTGTTTG GCTTTCGGTC	1980
	TAGGTAGCCG AGATGCCaTT CTCTAAGTTG TTTTAACACT TCTGGAATTA TCAGTACTGC	2040
50	CAATACTTGA TGTTCTAGAA GTGTTTTTAT TATGTCTAGC ATGAGGCTTT TCACCTCCTT	2100
	ACACATAATT TGTAAGTCAT CAACTAACCT ACAAATATAA TTATACTAAA CAAATGTTTA	2160

55

	GTTATCTACA	TTTAAATCTT	GAGAGAAATG	TTAAAAAGTT	CTAGTAAAAT	AATAGCACAT	2280
	TTTATCTTTA	AATGTAAATA	GAAAGCAGGT	ATGTAACGCA	CCTGCTTAAA	TAGaCATGAC	2340
5	TATGTCATT	TAAGTATT	CTCCCCATAA	GTCACCTAAT	ATCTGATTAG	GTGGGGCAGA	2400
	ACCATCCAT	GTTCTAATAG	GCAAGTAATA	ACGTTGCCCC	TCCCATGTAT	ATCCTACCCA	2460
	AACATGACCA	TCTTGTAACA	TCACTTCTGT	ATAATCACAA	TACCCACCAG	GTTGGAACTG	2520
10	ATAACCCACT	GGACAAGATA	AGAATGGCCC	CACCTTTCTT	ACTGTGATTG	GTTGATTGCC	2580
	GTTTGTGAAT	CTAGCACTTT	CTTCCATGTA	GTAAGTACCA	TATTTATTAC	GTTTCCATGC	2640
	ACTTGCAACT	GGTTTAACTG	TATTACTTGA	AGCGCTTGAC	TCATTAGAGA	CAGTGGCAAC	2700
15	CGGTATTTTA	CCATCCATGT	ACGCCCTAAT	CTGCTTGATA	AAGTAGTCTT	TAAGTTGCAA	2760
	CCGCTTGTCT	TCTGGCAATA	GACCGCGAGT	TACTGGGTCA	AAACCAGTGT	GTAAAACCGA	2820
20	ACTTCTATGA	GGGCATGATG	TTGAAGTAAA	TTCAATTGTGC	AATCTGATTG	TATTTCTGTT	2880
	TGCTGGTAAT	CCCCATTTTT	TCAACAATCT	AGCGCATTCT	TGGAAAGTTG	CCTGTTCAAT	2940
	TTTTAAGAAT	GTGCGTTAT	CTGCGCCCAT	TGATTGACAT	ACTTCAATAC	CGTAATAATA	3000
25	TTTATTACCT	ATTTGATTAG	CGGTATGCCA	ACCTACTTGT	GATTCATCTA	AGGCTTGCCA	3060
	AAGTGTGTG	CCTGATACGT	AAGTATGCGC	AATGCCCCGT	TCTAATCTTG	ATAAAGGTGC	3120
	ATTTACTAAT	CCGTTACGAT	ATGCTTCAGC	AGTCGCCCT	TTGCTCCCTG	CGTCGTTGTG	3180
30	TATAACTATA	CCTTTAGGGT	TACTACCACG	CTTAGGTAGG	TCATAACCTT	TAACCACATC	3240
	TTTGATGATT	TTAAGTTCTA	CTGCTTTAGG	TTGTGGCTTA	GCTGTTTCTT	TTTTAGGTGC	3300
	TTGTGTAGGA	GATTGAACTG	ATCGTGGCGC	TGTCTCACTT	TTAAAATTCG	GACGGATAAA	3360
35	CCACATAGGG	AAATCATAAG	CATGTTGTG	TCTTGTAAC	TTTTCCCAAC	CCCAGCCGGG	3420
	TTGTTCGATT	CCGTCAGTCC	AGCCACCGCC	TAGCCAATTC	TGCTCATATA	CAATGATGTA	3480
40	ATCTAAAGTT	GCTTCAATTA	CCCATGCAAC	GTGACCATAT	CCAGCACCGT	AGTTGCTACC	3540
	GAATACCACC	ATGTCGCCAG	GTTGTGCTAA	GAAGTCCGGT	GTATTTTGGT	ATACAGTAGC	3600
	TAATCCGTCG	AAGTTGTTAG	CGAACGGAAT	ATCTTTTGCA	CCTAAACCTT	TTAGAAGTAA	3660
45	TCCAAACAAA	ACTTTCCAAC	CAGCATTTGGC	ATAATCAAAG	CATTGAAATC	CATACCATAA	3720
	GTCCACATTG	AATTGTTTTT	CCTCAGAAGT	TTTCAACCAC	TCTATAAACT	CATTTTTAGT	3780
	TAATTTTGCT	TGCATTGTG	CCACCTCCAT	GATGATACTC	ATTCACATCA	AAGCCAACAT	3840
50	CGTTAGAGGC	GTCTGTGAAA	GGTTGTGATG	TATCATATTC	TTTTGGTGcT	TTCGCGCTTA	3900
	ATTCCGGCGT	TAAACTACTG	TCTTGATGATG	ATTTCCACGT	AAGTTGTTGT	TCTTCTTTTT	3960

55

	TTGGGTCAGT AATAACGCCA ATACCTGTAA GTAACGTGAG GATAGCGCCT ATAATTGCGC	4080
	TAGCTTGATT TAATTGAGTA GATAAATCTA ATCCGAATAA ATCCGTGACT TGCTTGATAA	4140
5	ATAGCAACAA TGCTCCAACCT AAACCAGTTA GTACTGCTTT GTTTTTGAAT CTCAATTTCC	4200
	AGTTAATATC CATTGTGTTG CTCCTTTTAT CCAAAATAAA AAAACGACTA AAAATTAGTC	4260
	GTTTAAAATT ATTCAATGGT CAATGTCGGA GATCCTGAAT AAACATCACT TATAGTGACG	4320
10	TACAACATCC CTGAAGGATT ACTAAAGTTG ATATTTTTAC TTGCAACTCC GCTATTGACT	4380
	CCTGATATTC CTAAATCACT TGACCCTAAA TTAGTTTGCG AAATCCTCAT TATACCGCTA	4440
	CGTACATTTT CTATTGTCAC CTGATAACTT TTATTGGGTT CAACTCCATT TATTGTCCAT	4500
15	TTTGCTGTTG ATTCTTCTAT GCTATCCGGA TATTTATTTT TAGGTAAGGG TTTTATTACA	4560
	AAAGATGAAG GCTTTTTCCA TACTTGGATA TTTCCAGCAT ATACTTTTGT ATATTCTTCA	4620
	CCTTCGTAAA TAAACTTCTT TACATTTTTA AAATTACCTT CCATAAAAAT CACCCTTTAA	4680
20	TTAAATATAA CGTATTCGGG TCTTTTGTAT ATATATAGTT ATATTCATTT TCTGTTCTG	4740
	TCCAAATTTT AACCGTCGGT TGAGATGCGC TTTTGTAGTT ATATAAATTA TCCGCTTGTT	4800
25	GTTTAGTAAA AGCTTGAGAT GACAAAACAT ACCGCTCGTC ATGATTATGA TTTTTTGGAG	4860
	CATATAAATC ATTTAGTGTT TGTTTGAATT CCTCAAAATC TTCTGTATTA ACTTTTGAGC	4920
	CAATCTGTTG CAATACACTT TCTGAAATAG AGTTGTTTTG TATTGCTTCT GCTAATTCTC	4980
30	TTAATGTGTT CATAGATTCA GGCGCGCTAT CAACTAGTTC AGCAATTTTT GTATCCGTAT	5040
	ACGTTTTAGA GTCGTTGAGA GTTGATCTT TGATTTTTTC AACTTCTTGC AATTTATTTT	5100
	CTAACCCCTC AACATTTGCG ATATTGATTT TGTTCAATAA CTCAGGTTCT GCTTTGATAT	5160
35	CTGTATCTTT ACCATCAATT TGCCACATTT TAGTGTCAGG ATTGATTGAT ACTACAGTAC	5220
	CGTTTACC GGGTGCGCCT TGTCTCCTT TTTTACCTGC TTCACCTTTT GCTCCAGGTT	5280
	GTCCCGGTTT ACCTTTATCA CCTTTCGCAC CTTTAAATCT ACTTTCATTC TTTTCGATGT	5340
40	AAGAAATGAC ATCTTTATCT ATTTTCTCTT TAAAGTCTTT GCTCAATAAA TCTGTCGCGT	5400
	TATCTTTTAA AATTCTCGTA ATAGCATCAT CTACCAATTT AACATCGATT TCTTTTGCTA	5460
45	CAGCAGATTC AATACCACTA TCAACGATAT TGAAAGAAAA GTTTGCGACA TGTATTTTTT	5520
	CTTCTTCTTT CTCTAAAAAC AGCTTACAGC GAACATAACC AGCGTGTGTTG ATAACCTTTT	5580
	TAGGTATCTT GTAGGTAAGG AAACCTTTTA CAACATCGTC GATAATAAGG GGCTCATTTT	5640
50	TGAATATAGA GCCATCTTCC ATAAACAAAT GTAATCTAGG TGTTAAGCCA TGTGCTTTTA	5700
	GATCGATACG ACCTTGTTTG TCATTGATAC CTATTCTTAT AGATGCTGTA TTTTCATCTT	5760

55

	CAACATCTTT TATTTTGTAC ATTTACACAC CTCCTTATTT ATATTTATCC CTTGTGAAGT	5880
	AGATACCTTT TAAGCCGATT TGTTTATATA ACTTAGCGAT TGTACTTGCT TGATGTTGGC	5940
5	ACCACTCTAT AGCAGTAGCG TATTGGTGGG TAGCTGGATT CTTAGGATTC CATCTAATTC	6000
	GGTACAATGT GTTTGTGACCT TTATTGATGT AATCCTTTCT TACGAAGCTA GCACCGCCCA	6060
10	TGATTGCTTT TGCTGGAGAT GTCCAACCTT TATTCCTTGC AAACGTCATT GCGTAGTTAG	6120
	GATTGTTGTC GTAAGCGCCA ATGCCGAAGT AGTTGTATAC TCCATCTTTT CCGTTAGCGA	6180
	AGTTACTTGT TCCATATCCA CTTTCTAAGA AAGCATGCGC GATTAAATAA ATTTCAATTAA	6240
15	TGTTGTGCTT TTTACAAGCT TCTGCGAAGC CTTTACCTTG ATTATTCAAT GTTCCCTTAC	6300
	CTTTAAGTAT CTTATTAAGT GCGCTAACTG AAACACCTTG ATACTTGCCT AAATTAAGCA	6360
	TTTGGTAGCA TTGTGTGTTA CTTTCCCAT A TACGCTTTAC ATTCATTGCT GAACTCGTTT	6420
20	GTGCTCGTGT AGCGTTAscC AACCCCAAGC ATTAGATTTT TTCGGGTAC CTCTTGCCAT	6480
	TTGTTTATCC AGTGCTTGTT TGAATGTATA AGGACTCGTT TCTGTTATGA TCTGCGGTTG	6540
	TTTAGATGCC GAACCATGTG TGGCTGTTGG TGACGAGTCT CTTACATTAG CTATATCAGC	6600
25	GTTTTTATTA TCTACCATAA CTTTATTCT AGATTTTGTT ACTGTTGGCT TAGTTATAGA	6660
	ATTTAATAAT TTTTCTCTGT TTTTAAATAT ATTAAGTAAT GCCTTTTCTA ATGCTTCGTA	6720
	TTTATCTTTA GGAGGAACAC CGTTGTCAAT CATATTCCAA TTAACATGTT CCAACATTGA	6780
30	ACGCCAAATG CTGTCGTCTA CTTTAAATT TTCAATACTT AGAGGTATCT CATATTTGGC	6840
	CATCATATCT ACAGCTACAA CCATGCGTG AATCTCATT AAAATAAATT CATTTTTACT	6900
	CGCACTATAA TCTTCACATA CGTCTATAAC TATATAATCA GGTTCAATTAG GAACTTCAAA	6960
35	TACAGCTCTT CTAGGTGCCC AAATATTATG TCTATCAACA TAAAAGTGGG GATATTCTAC	7020
	ATCCTGTTTG TATTTCTTCC TACTGTTATA TAAACTTTCT ACCGAGCTCA TCGTTTGTGC	7080
40	GTTTCTAATC ATTATTCCTT TAGGTTTTTC GAGTCGTCGA TTACCTTCTA CTATAAAGTG	7140
	ATAAATATAT TCTGGATAAT TAACCTCTTG GCTAGAAATA GTGTACTTTA TAGTTGTTAC	7200
	ATCTTTCCAA ATTGGAACCT TTTTATTATT TTTTTCGTTA TCATCACTAT CATCTTCTGG	7260
45	TTTAGGTGCC GGTGTAGTTT TGTCTGGATG ATATGGTGGT CTAACAAAAT ATTTAACCCC	7320
	TCCACCTGGT CCATCATGAT AAGAGTGTTC AATTTTATAA GGTGGACTTC CTGTTGCGTT	7380
	ATTTGTATAC CAGTTTTGAT CTACGCCATA CCAATAGTCT TTTGTGCATG GTCCCACTAC	7440
50	AATGTTTACA TGTCTGCCC AACCACAGT CCAAACACCC CAGTCGCCTG GTTGTGGTAC	7500
	AAAATCTTTT GTATTTCTAA TTATCTTGAA ATCTCTACCT CTATAATTGG ATTTTTGAGC	7560

55

	TAAATCCCAG CATTGTGCTC CCATTCCAGA ACCAGGTACA TCAATAGCTA TTTTGTTTT	7680
	AGCGATATAT AACGCCCATT CAACCACTTC ACTAGCTGTG GGCTTTCTAT TTTTCGGATT	7740
5	AGGTAATCCC ATGTATGCAC CTCATTTCAA TCAAAATAAA AAGCCAGTGC CGAAGCACTG	7800
	ACTCTTAACT GTTATTTACA TTTACCAAAC CAGAAGCACG CCCAGAAGCT ATATCCTAAA	7860
	ATCCCTTTAA GCATGGTAAT CACCTCCTTT AAATACCAA AACAGTTCTT AGTAAAGCTA	7920
10	TGACAATCGT ACTGAAGATA GTCCCTATCA AACCTAGAAT CCACATTTT ATGTCTCTAA	7980
	TATTCTTGGC ATTCTTTTCT TTATTCTTTT CATCTTCTAC CTTGTCGCGC TTTAATCTT	8040
	CAAAATTTCT ATCTAATTTG TCATAAATCT TTTCTTGCGC TCTAAGACTA TCTTCTATTC	8100
15	TGTCGAATTT TTCAAACATA GTCTTATCAT TTTCTTCTAA TCGCGTTAAA CGCCAATCTT	8160
	GTTTCATGTCG TTTGGTAAAT CCAAACATTA TGCCACCCAC TTTATTCAA TTA AAAAGCC	8220
20	ACAAGCATT CACCTGTGAC TTTTCATCTT TGTCTTCTGG ATATTTTCTT CCAGTGATTA	8280
	AAGCGTATTC TTCTTTATCG ATTAAACCCT TGTCTACGTA CCACTTAATT TGCTCGTTT	8340
	TATAGTAACC CCAAACATAA AAAGTTTTAA TGTCTTTAAA AGTTGGATAA ATCATCTTCA	8400
25	TTATTTAAAC GTCCCCCTCA GTACTTGTTC TGTAGTTT CAGTTCAGTC AACTGTTGTG	8460
	TTAACATAGC GTTTTGTTGA GCTAATTCCA TTGTTAATAC GTTTACTTGT GCCACCTGCA	8520
	TTTGCATACT CGCAACCATT CCGCGAAGTT CCTCATCACT TAAATCTGAC GCACTTTGTT	8580
30	GGTTTGATGC ATTCGGTACG TCTTCTTTT CGAAATTGCT ATTGTATTTA ATTTCCCGCT	8640
	TAGTGAAAAC AAACCTTCTA GGTTCGAACT CTTCTTTAAA TTTAATAGGC ACATTGTTAT	8700
	CATCTACATC TAAACTATTG CGTAAACCGC CAGTATTAAC GAATCCGATA ACTTCGTTT	8760
35	TATCGTTTAC TGTGATTTT ATTATTTCCA CCCATAATT TTAGTTATAG TAACTTTGTT	8820
	GGCAJTCGCT CCAGAACCTG ATGTTTTACC TAAATCAAAG TACACATCGT TATCTATTCT	8880
	TAAAGTAGTG CTAATTGTTT TGGATAGTAA GCACTCATAA ATACCGCCAC CGTTGCCGTC	8940
40	TGAGTCAACT ACATTGCTT TACTCAATTG AATCGCGTTA GGTAATGCGG TTAGTCCGAA	9000
	TCCCTCAATA ACGCCACCTG GATAAGTTCC ACTTACCAAC AAAATAGAAT AGTTTGTGTA	9060
	CGGTTTCAGTT AGATTGATTG TTGTACCTAC ACCATTTGCG CCACCGTCGA ACAATACCGT	9120
45	TGATTTATGT TCATTAGGAA CTGTCCACTG TTGCTCAAGT CTGCCGTTT TGATTGATCG	9180
	TGTGTAAATC TTTT TAGAGT TATAAGGTGT GAAGTTAAAT AGCTTGTTT TATCATCTT	9240
50	AACGAATACC GATAAATAAC CCTCATAACT TTCAACGCTA CCTGGTAAAT CCGGCACTCT	9300
	TGTTGCATAG TAATTACCAG CAGTTAAATA TCCCAAATCG CCTTGCGCAT TATTTAAGTT	9360

55

GAATTTATCA TCTACATACT GCTTAGCTTG ATTTAAAGCG TTGTTAGACG TTTCTTCAAC 9480
 AAATTGCTTA GTTAAGTTTC CATCATTCCT TTTATAAAAC GGGTACCATG TGCCGTAGAT 9540
 5 TTTGTATTTT GTGTACTCAT CGTTTGAATC GTCTGGGTAC CATGTTGCAC GAGCAGTATT 9600
 ATTATCAACA ACATAAACAA CTAACACACC AGATTTGCTT GATGTATAAG TTGATTCATC 9660
 GAACGAAGAA CCGTCATCAA CACCATCTTG TCCAGGCTTC TCTAACGTGC CTATATCCGT 9720
 10 CTTTTCTGGC GCATCTGTTG CATTAGTAAT ATGAATAATC CTAGATGTGT TAACTGCGCT 9780
 TAAACGCTA TCTATGGACT GCTCATACGA TTCAATGCT TTACCGTAAT CATCTGTAAG 9840
 TTTAGACTTT TGCCAATTCG TTGTTGAATT ACCTTTAACA AGGTCAGCGC CATTGATTTG 9900
 15 TTGTTCAACT TCGTTAACAC GTTCAAAAAT CGCTTGCTCT TTTTCAACTA TTTTATCGAA 9960
 TTCAGCTGTA ACAGCTTG TGCACTAGT TTGCGTCGCA GTAATAGCTT GTATAGCTTC 10020
 20 GTTTTGCTTG ATTTTCGATT GTTGAATGCC TTTTGTGCGA CTATCATTCA CTTTGTCTAT 10080
 TAACGTTTGT GTATCAGCCA TATTTTGCTT TAATTTGGTTA AAATCTTTAC CGACAGCTTC 10140
 GATAGTATCT TGAATAGATT TGATATAAAC AAGCTTTGTT ATACCATCAA ACCCACTAAC 10200
 25 TAAATCATTT TCAATATTGA AGCTAAATTG ACGTTCAACA ACAACATTAT TACTCCCGTT 10260
 TTGTGTAAAG AATGCCTGAG CATGCACCTT GCCTGAATGT TTTAAAAATT CATTGCGTAT 10320
 CACATACTGC AAACGCCCAT TAATTGCGTC TACTATCGTT AATTCGTCTG AAATATAAGC 10380
 30 GCCTCTATCT ACGTTATAAT CATCGGTTTT TAAnACGATA GATGTTTTAA CATGTTTCTA 10440
 ACTTATAGAT AAGGGTCTGT TATnCTTAGT 10470

(2) INFORMATION FOR SEQ ID NO: 21:

35 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3647 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

45 ATCAGATCTT GAGAATCGAG TTATTAAGTC TATCGAAGAC TTAATAAAA TCCAACCATT 60
 CATGCCTACA CAAGATTTTG ATTTTAAAAC TAAAGAAATT CAATCAAACA TTTCTGAAGA 120
 AAGATTTATC GAAATGATTC AGTATTTCAA AGAGAAAATA ACAGAAGGGG ATATGTTCCA 180
 50 AGTTGTGCCA TCAAGAATTT ACAAATATGC GcATCATGCT AGTCAGCATT TAAATCAACT 240
 TTCGTTTCAA CTGTATCAAA ATTTAAAACG ACAAACCCA AGTCCATATA TGTATTATCT 300

55

	TCAAATTGTA ACAACTAATC CTATTGCAGG TACGATTCAA CGTGGTGAGA CGACACAAAT	420
	AGATAATGAG AATATGAAAC AACTACTTAA TGATCCAAAA GAATGCAGCG AACATCGTAT	480
5	GCTAGTTGAT TTAGGACGTA ATGATATTCA TAGAGTAAGT AAAATCGGTA CCTCAAAAAAT	540
	TACTAAATTA ATGGTTATTG AAAAATATGA ACATGTTATG CATATCGTAA GTGAAGTCAC	600
	AGGTAAAATA AATCAAAATT TATCGCCAAT GACAGTTATT GCGAATTTAT TACCAACAGG	660
10	TACCGTTTCA GGTGCACCAA AATTACGTGC AATTGAAAGA ATATATGAAC AATATCCACA	720
	TAAACGGGGC GTTTATAGTG GTGGTGTGG ATACATAAAT TGTAATCATA ACTTAGATTT	780
15	TGCATTAGCA ATTCGAACGA TGATGATAGA TGAGCAGTAT ATCAACGTAG AAGCTGGTTG	840
	TGGCGTTGTA TATGATTCTA TTCCTGAAAA AGAACTGAAT GAAACGAAAT TGAAAGCTAA	900
	AAGCTTATTG GAGGTGAGCC CATGATCTTA GTTGTAGATA ATTATGATTC CTTTACATAT	960
20	AACCTAGTGG ATATTGTTGC TCAACATACT GACGTCATTG TTCAATACCC TGATGATGAT	1020
	AATGTGCTGA ATCAATCGGT GGACGCTGTT ATTATATCTC CTGGTCCAGG GCATCCATTA	1080
	GACGATCAAC AGTTAATGAA AATCATATCA ACCTATCAAC ACAAACCCAT TTTAGGTATT	1140
25	TGTTTAGGGG CTCAGGCACT GACTTGTTAC TACGGTGGAG AAGTCATTAA AGGCGACAAG	1200
	GTTATGCACG GCAAAGTTGA TACACTAAAG GTTATATCGC ATCATCAACA TCTGTTATAT	1260
	CAAGATATAC CAGAACAGTT TTCAATTATG AGATATCATT CATTAAATAAG TAACCCTGAC	1320
30	AATTTTCCAG AAGAATTGAA AATTACTGGA CGTACCAAAG ATTGTATACA GTCATTGAG	1380
	CATAAAGAAA GACCGCATTG TGGTATTCAG TACCATCCTG AATCATTGTC TACAGACTAT	1440
	GGTGTCAAAA TAATTACAAA TTTCATTAAAT CTAGTGAAGG AAGGATGAAA ACCATGACAT	1500
35	TACTAACAAG AATAAAAACT GAAACTATAT TACTTGAAAG CGACATTAAA GAGCTAATCG	1560
	ATATACTTAT TTCTCCTAGT ATTGGAAGT ATATTAAATA TGAATTACTT AGTTCCTATT	1620
	CGGAGCGAGA AATCCAACAA CAAGAATTAA CATATATTGT ACGTAGCTTA ATTAATACAA	1680
40	TGTATCCACA TCAACCATGT TATGAAGGGG CTATGTGTGT GTGCGGCACA GGTGGTGACA	1740
	AGTCAAATAG TTTCAACATT TCAACGACTG TTGCTTTTGT TGTAGCAAGT GCTGGcGTAA	1800
45	AAGTTATAAA ACATGGtAAT AAAAGTATTA CCTCaAATTC aGGTAGTACG GATTTGtTAA	1860
	ATCAAATGAA CATACAaCA ACAACTGTTG ATGATACACC TAACCAATTA AATGAnAAAAG	1920
	ACCTTGtATT CATTTGGTGCA aCTGAATCAT ATCCAATCAT GAAGTATATG CAACCAGTTA	1980
50	GAAAAATGAT TGGAAAGCCT ACAATATTAA ACCTTGtGGG TCCATTAAATT AATCCATATC	2040
	ACTTAACGTA TCAAATGGTA GGCCTCTTTG ATCCTACAAA GTTAAAGTTA GTTGCTAAAA	2100

55

AAGCAACACT ATCTGGTGAT AATTTGATAT ATGAATTGAC TGAAGATGGA GAAATCAAAA 2220
 ATTACACATT AAATGCGACT GATTATGGTT TGAAACATGC GCCGAATAGT GATTTTAAAG 2280
 5 GCGGTTCCACC TGAAGAAAAT TTAGCAATCT CCCTTAATAT CTTGAATGGT AAAGATCAGT 2340
 CAAGTCGACG TGATGTTGTC TTACTAAATG CGGGTTTAAG CCTTTATGTT GCAGAGAAAr 2400
 TGGATACCAT CGCAGAAGGC ATAGAACTTG CAACTACATT GATTGATAAT GGTGAAGCAT 2460
 10 TGGAAAAATA CCATCAAATG AGAGGTGAAT AATATGACGA TTTTATCAGA AATTGTTAAA 2520
 TATAAACAGT CACTTTTACA AAATGGCTAT TATCAAGACA AACTTAATAC CTTGAAAAGT 2580
 GTGAAGATTc AGAATAAAAA ATCTTTTATA AACGCAATTG AGAAAGAACC AAAGCTAGCA 2640
 15 ATTATTGCAG AAATTAAATC GAAGAGTCCT ACAGTTAATG ACTTACCTGA ACGAGATTTA 2700
 TCGCAACAAA TCTCAGATTA TGACCAATAT GGTGCAAATG CCGTGTCCAT TTAACTGAT 2760
 20 GAAAAGTACT TTGGTGGTAG TTTTGAAAGA TTACAAGCAT TGACGACAAA AACAACATTA 2820
 CCCGTATTAT GCAAAGACTT TATTATAGAC CCGCTTCAA TTAGTGTGc TAAACAAGCT 2880
 GGTGCATCTA TGATTTTATT GATCGTTAAC ATCTTATCTG ATAAACAATT GAAAGATTTA 2940
 25 TATAACTACG CTATATCGCA AAATCTAGAA GTGTTAGTTG AAGTACATGA TCGCCATGAA 3000
 TTAGAACGTG CCTATAAGGT TAATGCTAAA TTGATTGGTG TAAATAACAG GGACTTAAAA 3060
 CGATTTGTTA CAAATGTGGA ACATACAAAT ACTATTTTAG AAAATAAAAA AACAAATCAT 3120
 30 TATTATATTT CTGAAAGTGG TATTCACGAT GCATCTGATG TAAGAAAAAT CTTGCATAGT 3180
 GGTATCGATG GCTTACTAAT AGGTGAGGCG CTTATGCGTT GTGACAATCT ATCTGAATTT 3240
 TTACCACAAC TGAAAATGCA AAAGGTGAAG TCATGATGAA ATTGAAATTT TGTGGCTTTA 3300
 35 CATCAATAAA GGATGTTACA GCGGCCAGTC AATTACCTAT TGATGCGATA GGTTCATCC 3360
 ATTATGAAAA AAGTAAAAGG CATCAAACAA TTACCCAAAT AAAAAAGTTA GCGTCTGCTG 3420
 TTCCAAATCA TATCGATAAA GTATGTGTCA TGGTAAATCC TGATTTAACA ACAATTGAAC 3480
 40 ACGTATTAAG CAATACGTCA ATTAACACAA TACAGTTACA CgGCACAGAA TCTATTGATT 3540
 TTATACAGGA AATTAAAAAG AAATATTCAA GCATTAAAAT CACTAAAGCT TTAGCTGCaG 3600
 45 ATGgAAAACm TwATCCCAA caTtAAtnAA tnTTAgGGGG TCCGTGG 3647

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5966 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

	CcAcCTTGAC CACCTTTACG TGGAATCTTT TcMcCTkGAG CAACaTCGaT AATaTATATT	60
5	GAAAgTCAAC AAGTTCTGGA CTAAATGTTG CTGCTAAGTT ATCGCCACCA GATTCTATGA	120
	AAATTAGTTC TATATCGTCA TGACGTTCTA ATAATTCGTC TATTGCTGCA AAGTTCATAG	180
10	ATGCATCTTC ACGAATCGCA GTATGAGGAC ATCCACCAGT TTCAACACCA ATGATACGAC	240
	TTTCAGGTAG AACTCCTGAA TTTACTAATA TCTTTTCGTC TTCTTTTGTA TATATATCAT	300
	TTGTAATAAC GCCGATACTC ATTTCTTTTG AAAGACGTTT TACAACTTTT TCAATTAATT	360
15	GTGTTTTACC TGCACCTACA GGACCACCAA TACCAATTTT AATCGGATTT GCCACAATTA	420
	TAACCTCCTA TGATATGAAA tTCTAACATT GaCGTTCTCA TGCGCCATTT GATTTAGTTC	480
	TAAACCAGGC GCTGTCATGC CAAAATCTGC TTCTTTTAAT TCGAAAATCT GCTTTCTTGT	540
20	TCCTTCTATA TAAGGAATCA TGTGAGTAAC TATCTTTTGA CCAGCAGTTT GTCCAAGTGG	600
	AATAGCACGA ACAGCATTTT GAGTTAAACT TGAAACATTT TGATATAAAT AGTAATCAAT	660
	AATCGTTTCA ATATCTACAC CTAAATGATG GCCTAGCATA GTAAACAAA TAGCTGGATT	720
25	TnACTTTGCT TTCTTATCTT GCATTTGTTG ATGATACCAA GCAATCCATG GGCTATtATA	780
	AAGTTCTAAA GCCAATTTAA CCATGCGAGT CCCCATTGTG kTTGCACCAA CACGTGTTTC	840
30	TTTAGGTAAG TTTTGrACAr ACATCAGTTT ATCTATGTGT AATACTTTTT GTGTATCATC	900
	ATTTTCCAAT GCATCATAAA CTaaACGCAT GGCTAAACCA TCAGAATAGG TAAGTTGCTC	960
	TTGTAAAAAC ATTTTAAACC AAGCAATAAA AGTATGATCG TCATGAATTA TATTTGTTG	1020
35	AATATATGTT TCAAGACCAA ATGAATGACT GAAAGCACCT GTTGGAACCT GTGAATCACA	1080
	GAAC TGAAAT AATCTTAAGT GTGTATGATC AATCATGAGA ATGCCCTATA TGTCTGAAAG	1140
	CCTtATTAAC TTTACGGTCT TCTCGAACAT ATGGGATGCC TAAACTTTTT AATAAACTTT	1200
40	CAACTAAATA ATCATATTGT ACTAGCATTT CAGTCTCTGT AAATTGTGCT GGCAAATGAC	1260
	GATTTCTTAA TTGATGGGCT ATATCTCCCA TTTCTTGCAA TGTTCTTGGT TGAATCACTA	1320
45	AAAGATCTTC TGAATTAACA TCCACAATAA TCATATTATG GTCATCTGCG TATAAAATAT	1380
	CTCCATATTG TAAGTCAATA GGTGTGTTTT AACGAATGCC TATTTCACTG CCATGGTCTG	1440
	TAACGACTCT TTGAATACGT TTAACAAGAT CTGAATTTTC AAGGTATACT TTTTCGACGT	1500
50	GCTTTTGTTT TTCTGAATTT GACAAATTGG CAATATTGCC TTGGATTTCT TCAACAATCA	1560
	TTCTATGTTT CTCCTAGAAT AAGAAGTATC TTTGAGTTAA TGGTAACTCA GTTGCTGCAT	1620
	TACTTGTAAT TTTTCTCCA TCTACATATA CTTCATATGT TTGTGGATCA ACGTCTAATT	1680

55

	GACGCACCAT	GCGTTTTTAAA	TTTAATGCAC	GATTGATACC	ATTTTCATAA	GCAGTTTTAG	1800
	ACACGAATGT	CATTGACGTA	CTTGTAAGGT	TTCCGCCGTA	TTGACCATAC	ATTTTACGGT	1860
5	ACTTCATCGG	TTCAGATGTA	GGTATAGAAC	CATTTGCATC	GCCATTTACG	GCAGAGTTAA	1920
	TTAATCCGCC	CTTTACAAC	AATTCAGGTT	TAACCCCAA	GAAAATTGGG	TCCCATAAGA	1980
10	CAATGTCAGC	TAGTTTGCCC	GGCTCGATAG	ATCCTACATA	TTCAGAAATA	CCATGTGTAA	2040
	TTGCTGGGTT	AATTGTATAT	TTAGCGATAT	AACGTTTGAT	GCGATTATTA	TCATTATGTT	2100
	CAAAATCACC	ATCTAAAGGA	CCACGTTGTT	CTTTCATGCG	ATGTGCTACT	TGCCATGFTC	2160
15	GTGTAATTAC	TTCACCTACA	CGGCCCATTG	CTTGTGAATC	GGAACATAATC	ATACTGAATA	2220
	CACCCATATC	TTGCAGAACA	TCTTCTGCTG	CAATCGTTTT	TTTACGAATA	CGTGAATCTG	2280
	CGAATGCGAT	ATCTTCAGGA	ATAGCCGCAT	TTAAATGGTG	AGTAATCATT	ACCATATCTA	2340
20	AATGTTTCATC	TACAGTATTA	TGTGTATAAG	GCAAAGTTGG	ATTTGTAGAT	GAAGGTAAAA	2400
	TATTTGAAAA	TGCAGCGGAT	TTAATTAAAT	CAGGCGCATG	ACCGCCACCA	GCACCTTCAG	2460
25	TATGGTACAT	ATGAAGTACA	CGGTCTTTAA	CAGCAGCCAT	TGTGTCTTCC	ATAAATCCTG	2520
	CTTCATTTAA	AGTATCTGCA	TGTAATGCAA	TTTGAACATC	AAATTCATCA	GCAACATCTA	2580
	ATGCATGACT	CAAAGCAGAT	GGTGTTGCAC	CCCAGTCTTC	ATGTACTTTT	AATCCAATTG	2640
30	CTCCGGCATT	GATTTGTTCA	ATGAGTGCAG	TTGGATTTGT	TGCTTGTCTT	TTACCTGTAA	2700
	AACCGACATT	AATCGGTAAA	CcTTCCGGCAG	CTTCTAACAT	TCTATGAATA	TGCCATGGAC	2760
	CTGGAGTTAC	AGTTGTTGCT	TTAGAACCTT	CTGAAGCACC	AGTACCACCA	CCAATATGAG	2820
35	TCGTAATACC	ACTTTCTAAT	GCGACCTCTG	CTTGTTTCAGG	ATTAATAAAA	TGAACATGAG	2880
	TATCAATACC	ACCAGCAGTG	ACGATTTTAC	CTTCAGCGGC	AATGATATCT	GTTGTTGAAC	2940
	CTATAATAAT	GTCGACATTA	TCCATTATAT	CTGGGTTGCC	GGCATTACCT	ATGGCGAAAA	3000
40	TATAACCATT	TTTAATGCCT	ATATCAGCTT	TAACCACTTT	ATCGTAATCG	ATAATAACGG	3060
	CATTAGAAAT	GACAAGGTCT	GCAACGTTCA	CGTCATCACG	TGTTACACGA	GGATTTTGCG	3120
45	CCATACCGTC	TCTAATAGAT	TTACCACCAC	CAAAAGTAGC	TTCTTCACCA	TAAACCGCAT	3180
	AGTCTTTTTT	TATTTGAGCA	AATAGATTCT	TATCACCTAA	ACGAATGGAA	TCTCCAACAG	3240
	TTGGACCGTA	TAAGCTCGTA	TATTGATTTT	GCGTCATTTT	AAAGCTCATG	ATCTTTTTTCC	3300
50	TCCTTTTTTT	TTACGTTTTT	CAGCACCGTT	ATCTCCGAAT	ACACCTGCAT	ATTCATCATT	3360
	TTCATCAGTT	GGGCGATAGA	CACGTGACTC	ATCGATAGGA	CCATTGACCA	TACCACGAAA	3420
55	ACCAAAAATT	TTACGTTTGC	CAGCATATTC	AACTAATTGA	ACTTCTTTTT	TATCCCCAGG	3480

	TTCGAAATCT	AATGCTGCAT	TTGCTTCATA	AAAATGAAAA	TGTGAGCCCA	CTTGAATTGG	3600
	TCGATCTCCT	GTATTTTCAA	CTTCGATAAC	TGTTTCAGGA	TGATGGTTAT	TAATTTCAAC	3660
5	CTCTGTACTT	TTTGTAATAA	TTTCTCCTGG	TATCATTTGA	CTGCCTCCTT	TAAACAATAG	3720
	GGTGATGTAC	TGTGATTAAC	TTAGTACCAT	CGGGGAACGT	AGCCTCGATT	TCGATATCTG	3780
10	TAATCATGTG	TTGCACACCA	TCCATGACAT	CTTCTTTGTT	TAGAATTTGT	CTACCATAAC	3840
	TCATTAACTC	TGCAACGGTC	TTACCATCGC	GTGCACCTTC	TAATAATTCA	TCGCTGATTA	3900
	AAGCTAATGC	CTCAGGATGA	TTTAGTTTCA	AACCACGTGC	TTTACGACGA	CGTGCAACTT	3960
15	CCGCCGCCAC	TACAATCATT	AATTTGTCTT	GCTCTCGTTG	TGTAAAATGC	AAATTTAAAC	4020
	CCCCAATTTT	ATATTAGATA	CaATTTACAA	AATTTATATT	AATCCTAATT	GTTGTGATAA	4080
	ACAAGTAATA	TACAAAGTTC	AATGTGTAAT	TAGAAAATTA	TATTTTTAGC	ATATCCGATA	4140
20	TTGAAGCAAA	CAATCTAATC	GAAAACAAAT	AGTGGAATAT	ATTTATGTAA	AAACCAAAAT	4200
	AGTTTTTAAT	ATAACTTTTC	ATAGAATAGT	AGTATATTAA	TGAGTAATGA	TTCAAAGGAA	4260
	AGGTGAAAGA	TTGAAGATA	ATAGATGTGC	TTTTGAAAAA	TATATCTCAG	GTTGTGTAA	4320
25	TTAGTAATAA	ATGGACAGGA	TTATTTATCT	TAATAGGATT	ATTTGTAGCC	GATTGGACAA	4380
	TTGGATTAGC	GGCTATTGTA	GGTAGCATCA	TCGCCTATAC	TTTGC GCGT	TTTATAAAAT	4440
30	ATAGTGAGGC	AGAGATTAAAT	GATGGGTTAG	CTGGATTTAA	TCCAGTGCTA	ACTGCCATTG	4500
	CGTTAACAAT	CTTTTTAGAT	AAGTCAGGAT	TAGATATTGT	TATAACAATG	ATAGCAACTT	4560
	TATTAACGTT	ACCAGTTGCT	GCTGCAGTGA	GAGAAGTTTT	AAGACCATAT	AAAGTTCCGA	4620
35	TGCTGACGAT	GCCTTTTGTC	ATTGTGACTT	GGTTTACAAT	TTTACTTTCA	GGACAGGTTA	4680
	AATTTGTAGA	TACATCGTTA	AAGTTAATGC	CTCAAAACAT	TGAAACGGTT	AATTTTAGCA	4740
	ACAATGATAG	AATaCATTTC	ATTCAGTCAT	TATTTGAAGG	ATTCAGTCAA	GTATTTATCG	4800
40	AAGCGAGTGT	AATTGGTGCC	GTATGTATTT	TAATCGGCAT	ATTGATAGCA	TCAAGAAAAG	4860
	CAACACTCTT	AGCTGTTATA	GCTAGTTTGT	TAAGCTTTAT	CATTGTAGCT	CTATTAGGTG	4920
45	GTAATTATGA	TGATATTAAT	CAGGGATTAT	TCGGTTATAA	CTTTGTATTA	ATGGCAATCG	4980
	CACTAGGATA	TACATTTAAA	ACAGCGATTA	ACCCTTATAT	TTCGACTTTT	TTAGGTGTGT	5040
	TATTAACAGT	AGTGGTGCAA	CTAGGTACAA	CAACATTGCT	TGAACCGTTT	GGCTTACCTG	5100
50	CATTAACATT	GCCATTTATT	ATCGTGACAT	GGATTTTATT	ATTTGCTGGT	ATTAAACATG	5160
	ACAAAGTAGA	TGCTTGATAG	TTAAATCAAA	CCTAATATTG	TTTGAATATC	ACCTTAAACT	5220
55	ATACAGCGAA	TTGTATAGTT	TAAGGTGTAT	TTTTATGGAT	AAAATTAAGT	GCATACTTAA	5280

GTGTAAACT AGGAATAAAT AATTTATATT GTGTGTTGTG TGGGGTGAAT AATATGAATG 5400
 ATATGGATAA TTCCTTTTTA ATAACAACGG AAATTCAAAG AAAATGGATT GAAAAATTCA 5460
 5 AAGTAATTAG AGATACATTT AAGGCTAAAG CTGAATATAA TGATCAACAT AGCCAATTTT 5520
 CATATAAAAA TATTGAATGG TTAATTAAAG AAGGTTATGG AAAATTAACG TTACCAAAAG 5580
 CATATGGTGG TGAAGGTGCG ACCATAGAAG ACATGGTTAT TTTGCAATCA TTTTtagGCG 5640
 10 AACTTGATGG TGCCACAGCA TTATCTATTG GTTGGCATGT GAGTGTCGTA GGACAAATTT 5700
 ATGAACAGAA ATTATGGTCT CAAGATATGT TGGAGCAATT TGCTGTTGAA ATTAATAATG 5760
 15 GTGCATTAGT TAATAGAGCA GTTAGTGAAG CTGAAATGGG TAGTCCAACA AGAGGGGGAA 5820
 GACCAAGTAC ACATGCTGTT AAAGCTGATG ATGGGTATAT TTTAAATGGT GTGAAGACAT 5880
 ATACATCAAT GAGTAAAGCA CTAACACATA TTATTGTTGC TGCTTATATA GAAGAATTAG 5940
 20 AAAGTGTGG TTTTCTTA GTAGAC 5966

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
 25 (A) LENGTH: 17310 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTGTGTCATC GCGAAATAGT TAGGGTCATT CATTAATCCT TTTGAACGTA TTTTCATCAA 60
 35 ATATAACAAT TTCATTAGTA AAGGGGACTT GTTCAAACCA GCTATAATAC AAAATAGACC 120
 TATAGTCACA CTGCTTATAA TATAAGAGGT AACGATCACT TTTTGTCTAT TACCTAACTT 180
 AAAGATGATC ATCCCTAAAT AGAAATAAAT GACTACAAAT GCATATTTAA CTGTAGATGC 240
 40 AAGAACTTCC TTAACCGTAA TAAATATCAA ATCATCAAAA AATaGCaAAC AArGCGTAAT 300
 AATCATACGA TATGTATACA AAATAATGAm AAAGTGTmAA AAATGATTG CCTTTAATAA 360
 ATGGTTAGCG AAAACAGTA AATAAACTAA TATTAGTAAT GTGATAAAGT CAGCTATAGA 420
 45 AACATTCACA CCGCAATAA CCGAAGATTG CTGAATAAAA ACCGCTAAAC CGATAAGTAA 480
 CAATGTTAGT AATTTACTAT TGTGTTGATT TTCCATTATA AACGCTCTCC ACTTCTTTAA 540
 50 TCATTTTCTC CTCAGTAAAA CATTCTAAAT AACGTTTTCT AGATTGATTA CTCATTTTGA 600
 TGTAATCACT GTCTATTAAA TATTTTCCA GGAAGTTAGC AATAGTTTCG GGTGGTTGT 660
 TCATCATACA TATACCATTA TCAGCTACTA ATTCTGAAAT ACCGCCAACA TGAAGGCTA 720

55

	TTATTAAAAT AAACGTATCG TATTGTGATA ATAAATGACT CGCATTAATG ACATTGCCCA	840
	AAAATGTGAC ATCATTTTCT AACCCAGCTT GTACAACTTG TTGCTGACAA TCATTTAATG	900
5	TAGGTCCATC GCCTATAAAT GTAAAATGCG CATGATTACT GTTATGTAAT TTCAATATCT	960
	CTATTGCCGC GATTAGATTT TGTGGCAATT TTGGATAAGC AAATCTTGCA ATCATAACAA	1020
10	ATTGATGCTT TGTCGGGGCA TTAATCTGTA AATCTTGTTT ATTAGGCAAC ATTCCAACATA	1080
	CTTCGCCAAT ATTGTTATGT GATTGGCTTT TTAGCGTTTG CTTAACAGCG GGAACATCTG	1140
	CAATACCATT ATGTATTGTG GTTAATTTCA ATCGATTAAA TCGATATTTT AACGCTAACT	1200
15	GTTTATCGAA ATCTGAAACA CAAATAATGC TATCTGTAAT AAGTGACATT AATTTTTTCGA	1260
	TAACTAAATA TAGAAATTTT TTAGCTGGTT TAACACCCTC TGTAAAAGCC CATCCATGTG	1320
	CAGTAAAAAC TATACGTGTG TCTTTTCGATT TCGAAATGAa CTtCGCAATT CGTCCGACCG	1380
20	TtCCAGCTTT GGAAGAATGT AAATGGATAA CATCAGGTTT AATTTTCGAG AATAACTGTG	1440
	CTAACACTTT GACAGCTAAA ATATCTTGTT TAAAGTCAAT TGGACCTACT AAATGTTTGA	1500
25	TAATAATTAC ATTAACCTT GCATCTAGTT GTTCAATCAT TGGTCCATGA TTGCCTACAA	1560
	TGACATAAAC ATCATTGTGT ACGCAAAAAT GGTGCGGAG TTGAATGAGA TGTGTTTGTG	1620
	CACCACCATT GTCTGCTTTA GTAATACAAT ATATAATTTT CAACTGTTAC AAACCCCTTT	1680
30	AATGCTATAC TTTCAATTTT TTAACATGGC TATCTCATCA GATGAATAGT ATTTATAGCC	1740
	ATGCAAATCA ATGATGGCAC ATATTTCTTA ATGCCATTTG ATACTGTCTC AAGGGATTCC	1800
	TCGTTATACT GTAACAATTG GTCACAATCT TAAAATATA ACTTTTATTT GAACTTATTA	1860
35	AGTAAATTAA GACTACCTTG AGCCTTCCCC TGTAATAACA ACCATCAATG TTCTAATTGA	1920
	TATATATAGT TCCATCATT AACTACCTTT ATGTATATAT TTCATGTCAT ATTTAGTTT	1980
	TTGTGCGGT GTTAAGTCAT ATCCACCTTG AATTTGCGCA AGTCCTGTTA ACCCTGGTGT	2040
40	AACAAGACAT CTTTGCTCGA AACCTATCAC TTCTGAACTA AATAATTCTA CAAATTCCGG	2100
	ACGTTCCGGG CGTGGTCCAA TAAAACTCAT TTCCCCTTTA ACAACATTAA TTAGTTGTGG	2160
45	TAATTCATCA ATGCGTGT TTACGAATAAA CTTCCCGACA TTTGTTATAC GATCATCATC	2220
	TTTATCAGCC CATTGCGCAC CGTTTTTCTC TCGTTTTTTG CACATCGAAC GTAATTTGTA	2280
	TATTTTAATT AATTTACCCA TCTTCCCAAC TCTAACCTGA CTATAAATAG GGTTTCCTGG	2340
50	CGAATCTATG ACGATAGCAA TGGCGAATAT AACCATAATC GGTAAAGTTA AAAATAATAA	2400
	AACAATGCTT AAAATTAAAGT CAATCGCACG TTTAATTGGG TAATAGCTTT TTCTCACTTC	2460
55	TTCTAGTTTG TCTAATTTTC TTTGATAGGC ATAACCCTTA TTATTATGGA CAGCTTCAAT	2520

	AATTAAAGTA ATCCTTTTAAA CCTGTTTCTA CTGTATATTT AGGAACAAAT CCTAATGCCT	2640
	TTAAGTTAGA AATATCTGCA TAAGAATGCT TAATATCTCC TTTTCGTGCT TCTTTAAATT	2700
5	CATGCTCGAC TGATTTTCCA TATAATTCAC CAATAATACG ATAAACCTCT AATAAATTAG	2760
	TAAAGTGCC TGTACCAATG TTATAACCGT GTCCAATTGC ATCTTTGTGT TCCATAATTA	2820
10	AGCGTACAGA TTGAACAACA TCATATACAT ATACAAAATC TCTAGTTTGC AGTCCGTCAC	2880
	CAAAAATGT AAATGGCTTG TTATGCTCAA ATGAATCGAA CATCTTTGAA ATCACACCTG	2940
	AATATTGTGA CTTAGGATCC TGTCTTGGCC CAAATACATT AAAAAATTTA ACAACCGCTG	3000
15	TTGGTATGTT ATATAACGAA CAATAATTTA ATGTCGTCCG TTCGCCGTAA TATTTATCTA	3060
	TTGCATATGG TGATAATGGT AAGATTAATG ATTGATCACT TTTAGGCAAA TCAGGAAGAT	3120
	CACCATAAAC AGCTGCTGAC GAAGCAAAGA TAAAACGTTT TATATGATTA TTATATTTTT	3180
20	TAATGATTTT TAACAATCTT AATGTTGCTA CGACGTTTAT TTCTTGAGAT AAGATAGGTT	3240
	TCTCAACCGA CTCAGCAACA CTAAC TAATG CTGCTAAATG AATAACATAA TCAAATTGAT	3300
25	ATGTCTTCAT GATTTGTTCA ACTGCATCAT ATTCACGAAT ATCTAATTCA AACACATGAT	3360
	CGTCAGCCAA ACTTTTAATA TTTTCTCGTT TACCTGTTCT ATAGTTATCT AGAACATAAA	3420
	CATCATAATC TTGTTGTAAA TCATCTACTA AATGCGACCC AATAAAACCA GCCCCACCAG	3480
30	TTATCAAAAC TCTTTCCAAA TCTTCCACCT CATTTATACA TTAAAAATAT ATCATAAAAA	3540
	CATAAAGTAT TGTAAGCTTT TTATCGATAT TTTTATTTA TAAAAATAAA ATGAGATAAC	3600
	TTTGTGAATT TTTATTGAGA TAAATTAGAT AGTGGTGTTT TTGTGATGTT TTATAATATC	3660
35	TTGGGTGTGT TAATACTAAT AATGCTTTCA ACTGATGCAT TAGACTGTGA CATCATAACT	3720
	CACTTAAGAA CTTCGCTTAT TAATTTTCTA CCAATACACT CCCTTCTAAG TGCCTAAAA	3780
	AATCCTTACT GCTAAGTGAT TAACTTAAC AATAAGGATT TATTTATCAT TAGTGATGA	3840
40	TTATTAACGG AATCTCATAC CACCATCTAC AATAATTGTT TGTCCAGTAA TGTAATCAGA	3900
	GTCTTTACCA GCTAAGAAGC TCACTACATT TGAAACATCT TCTGGTTGAG AACTCTGCC	3960
45	CAAAGCAATC TGACTTGTA AATTGTTCCCA ACCCATGCT TCAGGTTTAC CTGCTTCTTC	4020
	GGCTGTTGCC ACTGCGATAC TTTCCATCAT TGGTGTTTGA ACGATACCAG GTGCGAATGC	4080
	ATTCACAGTA ATACCTTCAG ACGCTAAATC TTGTGCGGCT ACTTGTGTTA AACCTCGCAC	4140
50	TGCGAATTTT GTACTGCAAT ATAAAGACAA GCCTGGGTTA CCCTCAACGC CTGCTTGAGA	4200
	TGTTGCATTG ATAATTTTAC CGCCATGATT GAATTTTFTA AATTGTTTAT GTGCGGCTTG	4260
55	AATACCCCAT AGCACACCTG CAACGTTTAC GCCATATACT GTTTTAAACT GTTCTTCAGT	4320

GCCAAATTGC GCGGCAGTTT GTCTTAcTGC GTTAAATACA TCATCACGGT TTGATACATC 4440
 TGCTTTGATA GCAATAGCTT TTGTACCATC ACTTGATAAT TTAAGTGCAG CTGCTTTTGC 4500
 5 CCCTTCTTCA TTGAAATCAA CAACTGCTAC TTTGAAACCA TCTTCCACTA AACGTTCTGC 4560
 AATTTTAAAA CCAATCCCTT GTGcTCCGCC AGTTACTAAT GCTACTTTGT TGTTTGTCAT 4620
 10 AAAGATCACT CCTCAAATTT CTTTCCTTTA ATTACATTTT ACTCCTCTTC ATTTGAATAG 4680
 TACAACAAAG GTAGCTCCAT TTAACAAAAT ATTCAGATAT TTAAGGTATA GTTAAACGCA 4740
 CTACCATTAG TGATTGGCAA TGCGTTTAAA TGTCGTTTTA AAAGTTCTTA TGTGAATAT 4800
 15 TATTTTTTTA AGTCTCTCGA TTAGTTTGTC ATCAATCTTT TTTGAGACA TGGTCTTTTG 4860
 ATTCAATAGG CGGTTCCGTG TTATCACTGA CAACTTTAGT TGTAGCTTCA TCTTTATGTA 4920
 TTTCTTCGTT AAATCCTTCA AGGTTTTTAG TCGTGGGATT TTTAACCTCA GGATGTTCCA 4980
 20 TCATGTCTTG ACTATCAAGT TCCTTTTTAC ACGTGTCTTT ATGTGATGCT TGATTGCGT 5040
 TCCCTTTACT TTTTGAATA GTGGTAGTAT CTGCTGCAGC TACTAATTTT TTTCTACCTA 5100
 AAATAGATAT GGCTGAAACA AACCAGAGTA TTGCAGATAC AAAGTTGCAT AATACTAAAG 5160
 25 CGATAATAGC CAATACAATT AATATGACAC CTTTTGAAAT CCTTTCTTTA AATAAGTCAG 5220
 ATGCCAATAC GATGACAGGT ACGATTGAAA GTATAATTAC AAATATAGAA ATTATTGCCG 5280
 ATATAACTAT TGTTACTATT AAATAATCAG CTCTGCTACC TGATAATAAA TAGAAAAGGC 5340
 30 CGAAAATTAG TCCATAGCAA ATTACAAACC CACATAAAGT TATAGCCATG AGTACTATAT 5400
 AAGCTATTTG AAAATATAAA CCTATCTTTA TGAATGATT TTCTACATTT TTTCCATGT 5460
 35 CTATTCCCCA TTTATTTAAA ATTTATACTT TACCTTAAAT ATTCTCTTTA TTCTTTAGTG 5520
 ATTTTATCTT TAGATTCAAA TTGATTCTCT GTACTTTCAA TATCAACTTT TTCATTTTCG 5580
 TCTGTCGATT CATCTTTTGA GTATTTATT CAAATCAGCA AAATACCACC AATCAGCCAT 5640
 40 AAAATTGACG AAAGGAAATT ATATAAACAC AGTGCAATAA TAGCATAAAC AATAAAAAGT 5700
 GCACCTCCGA TTACAGAGTA ACTTTCCATA TAAATCGCAG TAAAGATGGT TGGTAAAACA 5760
 GTGAAAAGAG CCAATATTAA TCCTAATAAA AAAATTGTTT CGTAATCAGA TCCTCCAGCA 5820
 45 ATATTAATAG ATATCATCCT AACAAAAACG AACTAAAAAT ATATTGAGC TACGATGCCT 5880
 ATCCAAATTG CTATTTTTC TATAATTGAG CTCATACTCA TTCCCATT TTAAAAATT 5940
 50 TATACTTTAC CTTAATATAC CTTATTTTAT TTAATTTTTA TATGCAAAT AAAAAAATGG 6000
 AGAACTTCAA TATTTATAAA ATATCAAAAG TTCTCCACAC TATATTGTTT TATTATATTT 6060
 TCGCTATCAA TACGCTAAAT CATCATATTT CCCTCAACAT CACAGTAAAA CTATTGCTCC 6120

55

TTCCAATTGC GCAGTTGTTC AACATCATCA TCTTGTTTAA GTAATGCCAG TGGTACTTGA 6240
 AGATTAAGAC ATCGTCCTGA AATATTAAAG CGTGTCACAC CTGCTGGCAC AGTTTCCCCT 6300
 5 TTATGAACAA CCGCTTCAAT TTCCTTATAA CTCAATGGCT GATACTTCAT GAGTACATCT 6360
 TGTGTGAGAAA GACAAGGATA TGTACCTTGT GCAATTCTCT CTACAGAACA ACAACCACTA 6420
 10 TAACTTGCGA CAACCTTTTC CCATACTTGA AAATGTGCTT CGCCTAAATC TTTTGTATAC 6480
 AAATATTGTT CTGTATCACC ATGACACATT GTAATAAATG GCGCTTCTTG TCTTGTCTCA 6540
 GTAGTCCATG GCAAGCGATG TTCTTGTTGT AACGTTTCCC ACCACACACC AAATGGAACT 6600
 15 TTATGTTGCC ATGTACTAAT TGAATATTGT GTTTCATGGA TTTCTTGCAC TGGAACTTTC 6660
 TTACATCCTA ACGCTTTCAA ACTTGTATAC CGATGCACAC CATCTATAAC CATATATCTA 6720
 CCATGTTGCA TCGCTGTCAC TAAAATAGGA TGACGTATAA AATCATCTGC TTCAATACTA 6780
 20 CTTTTCGTTT TTTCCAATCT TAAAGGTTTCG AATGTTTCGT GAAGATCAAT CTTATCTACT 6840
 GGTACCAATT TTAAATGTTT ATGAATATGA TTCAATAGTT ATTCATCCTC CTTTGTTTGT 6900
 GTTAAATAAA TAAATTCAGG ATGTGGATGG CTTAAGAAAT CGTGATGTGA AATAGACCAT 6960
 25 CCGTATGCAC CTGCATATTT GAAAACAATA ACGTCGCCCTG TACTGATTGC GTCTATCTGT 7020
 ACTTCTCTAG CAAAGACATC TTTCGGTGTA CATAATTGAC CGACTAACGT TGTGTCCTGT 7080
 CTCGAAATTG AAACTTTTTC AAATGAATAT GGATTGTCCT TATAGCGATA AATGTCAAAA 7140
 30 GGATGGTTAT GTTGCCAAGA TACCGGCAGT CTAAATTGTT GCGTACCTCC TCTTAATATG 7200
 GCATACCAAG CACCATGTAC TTTCTTAATG TCTAGCACTT CTGTCACATA GTAACCAATA 7260
 35 TGTGCCACAA TAAAGCGCCC ACATTCAAAG TTCAATGTCA CATCTTCCAT TTCTTGCTCA 7320
 ACGATAAGTG TTTTAAAACG TTCTACAAAA TTATCCCAAT CAAATTGGTT AGTTAAATCT 7380
 GCATAGTTAA CGCCTATGCC ACCACCAAGA TTGATATGTT TGAGTGGAAT TCGATGTTTT 7440
 40 TCAGACCATG CCTTTGCTTT TTTAAAATAA AGTTTCACTA CATCGACATG TAAATTCGAG 7500
 TCTAAATTGT TAGAAATAGA ATGAAAATGA AATCCATCTA GATGAATCTT TGGCATTGCG 7560
 AGCGCAGcTT CAATGACATC ATCAACTTCG TCTTCAGAAA TACCAAATTG TGTGGGCGT 7620
 45 CCTGCCATAT GCAACGTTGC ATTGGGAAAT GGTCTGCTA AATTAACACG CAATAAAATG 7680
 TGTGTGTCT TATCTTCATC TTCTAAGATG GCATTTAGCC GTTGTAATTC ATGCATACCT 7740
 50 TCAACATGAA TACGCTGAAC ACCTTCACTT ACTGCATATC TTAGTTCCTC GTCTGTCTTA 7800
 CCAGGGCCAC CAAAAATAAT ATGATTTGCT GGTTTAAAAG CAAGACCTTT TGCTATTTCA 7860
 CCTTGAGATG CAACCTCGAA TCCTTCAACA TACTGACTAA TTGTATCTAG GATTTTTCGT 7920

55

	TGTTGCAAAT GATGTTCCAG TCCGACTAAA TCATAGATAT AATGACAAAC TGGATGAGAT	8040
	TGTGCTTTTA ATTGTTCAAT AACAGGTTGA ACTATACGCA TTAGCCTTCA TCCCCTTTCT	8100
5	GTTTAGACGT CGCTAGAGAT GCACTTAAAT GCGGATATAT TTTCCGCGA TCATCACCTA	8160
	AAATAAATGT TTGTACACCT TGTGCCTGCC ATTTTGCAAT ATCTTCATCT TCACGTGGTA	8220
10	ATGCACAAAA ATGTTTACCA TGTGCATTCA CAACTTCAAA AATATGTTGA ACATGTGATG	8280
	TTACTTGATC ATCACGCGTT TGCCATGGTA TGCCAAGTGA CTGCGATAAA TCTGCGGCAC	8340
	CTTCGACTAT CATGTCTAAA CCTTCGACTT GTGCTATATC GTCAATGGCC ATAACCCCTT	8400
15	CAACATCTTC TATCATGGCA ATCACCATAA TATGCTCATT AGCCATCTCC ATTGCATCAA	8460
	GTAATGGTGT ACGTCCAAAT CTTGCCATGC GACCACCATT CAAACTTCTT AATCCTTGCG	8520
	GGTAATAACG ACTTAATTTT ACAATATGCT CAACTGTCTC ACGATCTTTA ACGTGTGGCA	8580
20	CAATAATACC TCTCGCACCC ATATCCAACA CTTTAATGAT ATCTCTATCT ATCACTGCAG	8640
	TGACACGTAC AATTGGTATA ATATGCGCTG CTTCAGCTGC ACGAATTAAA TCGGCTAGTG	8700
	TCTCATCATT AATCGCCACG TGTCTGTAT CAATCACAAC AAAGTCATAC CCGCTTGCTG	8760
25	CGATAACCTC GATCATCAAT GGGTCCGGTA TAGAATTAAA AATGCCATAA ACTGAATCAC	8820
	CATTGTTTAA TCTATGTTTC AGAGATAGTT GTTGCATCAT TGATACCTCC TACACCTAAT	8880
30	GGATTTGTAA CATGATGAAT TCTTAACTCG GAGTCACTTA ATAATCGACG TGTGTTAAC	8940
	TTTTCAACTT GAATCGTAGG TTCAAACAAA TCGAAATGTT GATAGTTATT CAACTCTGGA	9000
	AATGCTTCTT GATACGCCTC GATGATGCCT TTAACCCATT GCCATTGCAG CTCCTCATCG	9060
35	ATACCATATT GCTTTTCAAT AAATAAGATG ATTTCCGGCGA TATTAATAAA GAAAAATGCA	9120
	TCATGTAAAA AGTCGCGTAC TAAACGTTTC TCATCTGTTT CAATAAATGA ATTACTATTC	9180
	ACTTFTTTAT GTGCTTCTGG CATTGGCTTT AATGTCAGGT GTGAAGCAGC TTCACTTAAA	9240
40	TGctCACGCT TAAAACGAAC ACCATCATGG AAATCTTTTA AGGCAATACG TGTAGGCCAA	9300
	CCATTTTCAT GAATGAGCAT CATATTTTGT GCATGCGATT CAAAGGCAAT ACCGTGATAA	9360
	TAAAGCATAT GAATCATTGG ACGAATCGCT ACAGCTAAAA ATTGCTTTGT CCAAGCTTCA	9420
45	GAACCATATT GTTTAATCCA ATTTTCAATG AATGGTACAC CATCCTTATC ACTTGATAA	9480
	AGTGCAATTA ATGGTATCGC ATCCTCTTCA TCGATTAACA TATGATATAT ATTTTCACGC	9540
50	CATATAACAC CTAACGCACC ATAACTTGA GTTTGTTTAT AAGGCGAAAG TTGTGTATTT	9600
	AAATAAGACT GTCCTAAGAC TTCCCCTAGA AAAACTGTCT TTAATTCATC TTTTAAATAC	9660
	ATATCTTGTT GCTGTATCTG CTTTAACCAA TCCGTAATTT GCGCTGCATT TTCAATTGTA	9720
55		

	TATTTTGTGCG	TGTCTATTGG	CGACATCGTA	CGAATCGATT	GTTGAGGGTG	ATATAGCTCA	9840
	TCACTTTCCC	CTAACCATAG	TACTGTGCCA	TTAAGCCTTT	CTTCAGCCAA	ATCAACTTGG	9900
5	ATGACATGTT	CAAAC TGCCA	TGGGTGTACA	GGTATCATCT	CAACATCATT	TACATGTTTG	9960
	CCAGATGCTT	CAATTTGCTG	TACAAAATGT	TCATAAGTCT	TATCGCCAAC	TTGTTGACGT	10020
	AACATTTTCGT	TAACTACAAC	ATTTCTTGAT	ACCGTCGTTT	CTACTTTATC	TTTGTGCGATA	10080
10	GCTAACCCT	GCAGTTTAAC	GTTTGGTACA	AAATCAGGAC	CAAATTTCAA	ATTATCACTC	10140
	AACGTAAATC	CTAAACGTGA	TTTGTAACTT	GGATGATACT	GATGCCCTTC	CATCGCATAA	10200
15	AATTCATAGT	CGTTAAATGT	CTCAGGTGTT	GCTGGTGGGT	TTGATTCTCG	ATACTGCATA	10260
	CTTTGCGTAT	CTTTTAATTC	TGTCTGTAAT	AACTCGACAA	TAAATTGTTC	TAGCTTTTCA	10320
	TCATTTTTAG	GAAATGTAAA	TACAACCTCT	CTCAATAATT	GTGTATAGTC	TGTTGTTGTA	10380
20	TCTGCCTCAT	CTCCTACGAC	ACGCTCAATT	GGTGATGTGA	TACGTATACG	ATCAAAGCTA	10440
	TGTGTCTTTT	CAGCAGTAAA	ACGATACTCT	GAATCATGTC	CTTCTATTGT	AAAATGACCG	10500
	ACACCGTCTT	GATATGACGC	TTTATACACA	ACAATATTCT	CATAAATAAG	TGATGATACC	10560
25	AGTTGGTGCA	TCACTCTAGT	CTTTACACGA	TTAAGAATTG	TTTGATTAC	AATACGATAC	10620
	CTCCTTGTTA	TGACAAATTG	GATTTGGTAT	ATGTGTATAA	ATAGGGTTTG	CACCACAATC	10680
	ATTCAATTTA	CTCATCAAAT	TCGCTTTAGC	CGCAATGGTC	GGCGTTTGAT	ATAAATCTTC	10740
30	TACACAGTCA	ACAAATACTG	CGTTATTTCG	GTATTCTTTT	TTCCAAGTCA	TAAGACGATG	10800
	CGCTACAAGT	TGCCATAACA	CAACTTCATT	TCTAGTCGCT	TTACCAATAG	TTGATACTAA	10860
35	ATGTCCTAAG	TGATTTACTA	CAACGTAATA	TTTAAGACGA	TGCCATGCTT	CATCATGTGC	10920
	ATATACAACA	GGGCTTGATG	CTGCCACAAC	ATTTGGCACA	AGCTGTTTTT	CAGTAGCAAT	10980
	CGTTCTAGAT	AGACAAATGC	CTTCAAGATC	TCTGACAAAG	CATACGTCGG	GTATGCCATC	11040
40	TTTTAATTCA	ATTAATGTAT	TTTGTACATG	TGCTTCTAGA	CTAATGCCTG	TGTTACTAAA	11100
	CAGCTTTAAT	ATCGGCAATA	ATGTACGATT	CAAATAACAT	TCAAGCCATG	CTTCTGGTGC	11160
	TAAACCACTT	TGCTCAATCA	CTTGTGATAA	CTTAGACATC	GGTGAATCAG	GCATCGTTTC	11220
45	AAATAATGAC	GCCAATACAT	GAATATCTTT	ATCAGCATGG	TAATTCGGTA	TCCCTTCACG	11280
	AACAATCATG	GCACTATTTG	TTAATAAATC	CATTTAGGGT	TCAACTGTTT	GCCCTAATGG	11340
50	ATTCGGTAAC	AATGCACGAT	ATCCTTCTTC	AAACATCAAT	TTAAAATGGG	GTGTTTCAAC	11400
	CTCATCTTTG	ACTGATGCGA	TAACTGCGC	GGCATCAATT	GTCCGTTCAA	TCTGTTCAAG	11460
	GTCATTCGTA	CGTATAAAAT	TAGTGATTTT	AACGTGTATC	GGTAATTTTA	AATAAATGTT	11520

55

	GCCAAGGTCT	TTTATTAAAC	CTTGTTCACT	ATATTGCATA	TACTGTGGAT	GCTGTCGCAA	11640
	CACATTGATT	TGATAAGGAT	GTGTTGGTAA	TAAAATAAAA	TCTTTGGGTA	TCTCTGATAT	11700
5	ATCTATGTCT	GCTAATTGAT	ACAACACTTT	CTCAACCTGA	TCTTCTTTAC	CTTCTACATA	11760
	GCGCGTGAGC	AGAACATCTT	GATGCACAGC	TAAATAATGC	AATTGGAATG	ATGTATGACA	11820
10	TTCGGGTGCA	TATTTCTCTA	AATCTGCTTC	TGAAAACCCA	CTTGCACTCT	TAGGAGTCGG	11880
	ATGAAATGGA	TGACCTAAGT	ATAAAGATTG	TTCTGAAACG	ATATAACGAT	CCTCTACGTA	11940
	GTCTATTGTG	TTACTTTGCA	AATAACGTGC	CGTGCGATGA	ATGCTATTAT	CGATGTCAGA	12000
15	CATAATTTGC	GCCATATGTT	GTTGCACTGC	CGTTTGATTA	TCTGCACTTT	GAGCCATATG	12060
	TTGCAAAATA	CGCGCAATTG	CTTCTTTATA	AGTTGTTATT	TTTTTACTTT	TTCCATCGAT	12120
	AAGCCATACC	TCTGGATGAT	ACATATGATG	CCCCATCGCA	GACCAATAGC	GAAATTCACC	12180
20	CGTTAAAGTT	TCGAGCTCTG	ATAATTGTAT	AGACCATTGA	TGATTTTGAG	GTGGTACTTG	12240
	ATATAAATTT	TCTTCTCTAA	AATATTCATT	TAAAATGCGT	TCGATAGCCG	CATACGCTGC	12300
	ATGTTGTATT	AATTCTTTAT	TTTGCACTTT	TTTGTTCCTA	CTCCCATAA	TTCATTAATG	12360
25	TGTGATCGTT	GATTTGATTA	GTGATGGTTG	AACAAATTAA	AAATAAACTA	CTTACTGCAA	12420
	ATACTACGCC	CATAACGATA	AACGTAGTAG	CTGGTGTAGT	ATAACTTGTA	ATGGCAGCGC	12480
30	cACTaAGACT	GCCAATAATT	TGACCAACAA	CTAACATACT	GTTTCGTCGTT	CCAACAAATG	12540
	TGCCTTTAAG	TTGTTGATGA	CACGCATTCA	CGACAACAAA	CATGACACTT	TGAATCAATG	12600
	CACTATATGT	TAATCCTTGA	AGTATTCTTG	CAGCCATTAA	AAACTCTATA	TTCGTCGCTA	12660
35	AACCTTGCAG	TATCGCACTA	CAACCACATG	CAATCGTGGC	AAATATATAT	ACTGATTTAA	12720
	CATATGATTT	ATCATTAAAG	CGTCCCCATA	AAGGCGCGCT	TAATATCGAA	GCCGTCCAAA	12780
	ATGCGGACTG	TAAAAATCCA	ATCACACTAC	GGTCATCTAT	CGCTGTATGA	TTCCTGATG	12840
40	AAGCAAGTGG	TGATAATGCA	GTTAGCATGC	CATACATAGC	AAAGTTTGCT	AAAACGCCAA	12900
	CGATAATAAA	TCGACATGTT	TGTTGTGTGC	ATAATAGACA	TTGAAATGAA	CGGCGAATAC	12960
45	CTTTATTAAT	ATTTGGTGTT	TGTGATTTTG	GCATATGTGT	CGTTTCAATC	AATTTTAATG	13020
	CACCGAAAAT	ACAGACAATA	AAAGTAATAA	CGGCAATACT	CATCAGTAAC	GCACTAAAAC	13080
	CTAATATCGA	AGCTGTAACA	CCGCCAATTA	ATGGCCCCAC	AAGAGACCCT	GCGCTGACTG	13140
50	AACTTTGCAG	TCTTCCTAAT	ACCTTTCCAC	GATCTTCAGC	TGGCGCCTCT	GCACTCGCAA	13200
	ACGCACTTGA	TGCATCAACA	ACACCACCAA	ATAGTCCCTG	CAATAACCTC	ACAAGTACAA	13260
	ACTGTAATGG	TGTCGTACAC	AATGCCATTA	AAAATAAGCA	TACCGCCAAA	CCAAGTAACG	13320

	ctATCATCGT	CGTTACAGCT	GGAGCAGCAA	TCGCTATACC	ACTCCACAAC	TGTATTTCTA	13440
5	CGACTGATAG	ATTTTGTAGT	GATGCCATAT	AAATTGGCAA	TAATGGCACA	AGTACTGTCA	13500
	GTCCAGCAAT	CGCTATAAAC	TGACTGAGCC	ATAAAATGCG	AAAGTTACTG	CGCCATATAG	13560
	ACTGATTAAT	CATATGTCAC	CATTGGATTT	GGTACGGTAG	TTAAACCTGA	AGGCATACTA	13620
10	CCTCCACCAC	TATCACGTTG	ATATAGCAAT	GGTAATAAAA	TTTGTTTGAA	TGGCCACGTC	13680
	TGTTTATCAA	ATAAAATGTG	TCTGACAGCT	AGCTGATCAG	TTGTAACCCA	GGAAATAGTT	13740
	GCCACTTCAT	TTTTTAAAAAT	TTGTTTTAAC	AACGACATAA	GTTTCATGCTC	ACTTACACCA	13800
15	AATAAATCTT	GAATTGCATC	AATAATGGCA	TATAGATTTA	CCGATACAGC	TAATGTTTGA	13860
	AAATAAGCAA	AGAATGTTTC	CAAATCCTCA	TTAATTAGCG	TATTAGGTGT	ATCTTCTCTG	13920
	ACGACATACT	TCGGCAATGA	AAGCTGATGT	GCTGTTAGCC	ATGGTTTATA	AATTCTGACA	13980
20	GTATCATGAT	CACGTAACAC	GCATTTTTGT	ACACGTCCAT	CTTCAAATGA	CAACAATATA	14040
	TTTTGACCAT	GCAACTCTGG	TAATGCGCCG	TATTGCATAA	ATGATAGTGT	TACCTTTAAA	14100
25	AAGACTTGCG	CGATATCTTC	AAATAACGTC	ATGACATCAT	TTTTAGAAAT	ATTATCTTTT	14160
	CCACAAATCA	TTTGATATAA	AGTGCGATCA	TTTGCCGCGA	GTGCTGCCAT	TGACACTAGC	14220
	TGTTGCGTAT	CATTTTTGGC	TAGCACTTCG	GGATACTTTC	TTAGCTGAAC	AGTTAGATGA	14280
30	CCTAATTGAT	CTTTGAAAAT	ATCATTATCT	TGACCCATAT	ATGACCACCA	AGCTGTTTCA	14340
	TCACAAACCA	TGACATACTT	AGCTAGTGCT	TCATCTTTTT	CTATAAGCTG	ACGTAATAAT	14400
	TGTTCTGCTT	GTTCTCCGTT	TTTCATGTAA	CGCGTAGGCG	TTAGCCTTAA	TGCGCCTAAT	14460
35	GACTGCATTG	CAAATGGTAC	TTTGACATGG	TTATACGGTG	CGCCAATATC	AATTAATGAA	14520
	CGCATACTTG	AAGACGACAG	ATAATCTCCA	AATTTTAACG	GTAATAGTAC	AACCAACTTT	14580
	TCAC T AATCT	CTTTCGCAAA	GACGTTGCGC	AGAATATGCT	GATATTGCCA	AGGATGTACC	14640
40	GGAAATAGTA	CATAGTCATC	TATTGATAAC	CCTTGATCAT	TTAACATGTC	TGTCGCTTGT	14700
	TCTTTTATAG	GTA C TGTCAA	ATTTTCTAAT	TCATCGATAT	TTGCAGTATC	GCCATGAATC	14760
45	ATATGTGTCT	TTTAACTGTC	TGCAACCATT	AAAGGAAATG	ATTGATTTAA	TTCAGCTTGA	14820
	TACACTTGAT	AATCCGCTTC	TCTTAATCCT	CTTTTTTCTT	TAGCTAATGG	ATGAAATGGA	14880
	CGATCTTTTA	AACTTGCAAA	CTGCTCTGAC	ATCACAAAAG	GATGTGACGC	TAAATCTAAT	14940
50	TCTGATAATT	GTTTAGCAAG	CTGTGTGGCA	GCAGTAGTCA	GTCCTTCTTC	AACGCGAGCC	15000
	ACTTCCCAT	CATGACTTAG	ATCACAATTC	ATATTAGCAA	TTGTTTGCCA	AAATTCAGCT	15060
55	GCCGTTAAAG	GTTGCTTAGA	CACCCTTCCC	TCTATCGTAA	TTGGTTGTGA	ACTTTCGTAA	15120

	TATATCAAAA	GCGTTTGTCC	GTTTTCTTTA	GTAATCTCAC	TATTCGATAC	AATTCCGGCT	15240
	ATATCTTCAA	ATAATAATGC	ATCAACTAAA	TCTCTTAATA	TTATCGCTTG	TGCTGTATTG	15300
5	ACTGCTGTAT	GATTCTGCAA	TGTTCTAGACA	CCTCGCATTG	TTAATATAGG	TTCAATGTTG	15360
	TCCCAATATT	TTGTTGTTGT	GCCTGTTGAT	AAATAAAATA	AGCACTTGAA	ATATCTTCGA	15420
10	TAGCCATACC	CATCGGATTA	AGTAATATGA	TCTCATCATC	GTCTTCACGT	CCTGGTATGT	15480
	CACCTGTCAC	AAGTTGTCCT	AGTTCAGCAT	GAAGAGCTTC	TTTGCTGAAT	TTACCTTCTA	15540
	ACACCAATTG	GTTAATAGTT	TTCTTTTCTC	GATTACATTG	TGACCAGTCA	TCTACTACGA	15600
15	CTTTGTCAGC	TTTAATAAAG	ACTTCTTTAT	GCACATCCAT	GATAGAAATG	TTGCTAATAA	15660
	ATGCACCCTT	TTGTAACCAA	TCATATTCAA	TGTATGGTTG	ATCCGTTACG	GTACATGTAA	15720
	TGACTACTTC	ACCATTTGAT	ACTGCTTCTT	TAGCATTTTC	TGTCGCAATA	AAATTAATTT	15780
20	CCGGACGCTG	TTGTTGCCAT	CTATCAACAA	AGCGTGACAA	TGCTTCAGAG	AATTGATCGT	15840
	AAACAAACAC	GCGTTCAATA	TGATCGAATT	GCTCTAACAT	ACTTTGTAAT	TGCTTGTCTC	15900
25	CGATTAGCCC	GCATCCAATG	ATTGTTAAGT	CTTTAAATCC	TTTTTTAGCC	AAATGCTTTG	15960
	CTGCAATCAC	TGAAACTGCT	GCAGTACGCA	TACTACTAAT	TAAACTTGCT	TCCATAACTG	16020
	CAATTGGATA	ATTCGTTTCT	GGATCATTCA	AAATAATGAC	GCCACTTGCA	CGCTCCATAT	16080
30	TACGTTTCGA	TGGATTGTCG	TGCTTACTAC	CTATCCACTT	AATACCTGAA	ATTGCGTGTT	16140
	CACCACCGAT	ATGACTTGGC	ATTGCAATAA	TTCGATCTGC	GATGTGTCCA	TTTTCAGGAT	16200
	CCtGCTTAA	ATACGGCTTA	AGCGGTTGTA	CAAAATCATT	GTGCGCATGG	GCTGTTAATG	16260
35	CTTCTGTAA	TGCGTCCACA	TAAACTTGTTG	AATGATTACC	TCCCGCTTGT	TCAATATCTG	16320
	ATCTATTTAA	ATACAACATC	TCTCTatTCa	TTCTGaTTTA	ACTCCTTGTC	TTGATTTTCAT	16380
	TTTTTCTAAC	CATGTATCTG	AATAAACTAA	ATCTAAGTAA	CGATCGCCTC	GATCTGGTAA	16440
40	AATCGTGACA	ATTGTTGCAC	CTTCTTCAAT	TGACGTTATC	AACTGCTCAA	TCGCTGCAAT	16500
	AATCGAACCT	GTTGAACcTC	CGGCAAATAT	GCCTTCATAA	TCAATCAGTT	TTGACAGCC	16560
45	CAAAGCAGAT	TGATAATCAT	CTACATGGAT	CACTTGATTA	ATTTCTGATC	TATTCAATAT	16620
	TTGCGGTACA	CGACTAGCAC	CGATACCAGG	TAATTCTCTA	TTAATAGGTT	TGTCACCAAA	16680
	AATGACTGAC	CCTTTTCGCAT	CAACAGCAAC	AATTTGTGCG	TTTGATGCA	CTTCTTTTAT	16740
50	TTTTCTACTC	ATACCCATAA	TGCTACCTGT	CGTGCTGACT	GGCGCGACAA	AATAATCTAT	16800
	AGGTGCTTA	ATTGTTTCAA	CAATCTCTGT	GCCTGCACCA	TGATAATGGG	ATTGCCAATT	16860
55	TAACTCATTC	GCATATTGAT	TAATCCAATA	TGCATCGTCA	ATAGTGGCTA	ACAGTTCTTG	16920

TACATTGGCA CCATAACTTT TAATAATTTT CAAATTTGTT GGTGATATTT TAGGATCAAC 17040
 AACACACGTG AGTTTTAATC CCTTGATTTT AGCTATCATT GCCAACGCAA TGCCTAAATT 17100
 5 ACCAGAAGTA CTTTCAATTA AATGTGTATT CTCAGTGATT AAACCATGTT TAATACCATG 17160
 TTCAATGATG TACTTGGCAG GTCGATCTTT CATGCTGCCT CCAGGATTCA TATACTCTAA 17220
 10 CTTTGCAAAC ACTTCAATGTT TCGGAAATAG TTGATGAAGT TGAACCATAG GTGTTTGCCC 17280
 TACAGAATCT AACAATGAAT CGTGCACATG 17310

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5423 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ATACTAGTAA GCGCATCGGT TATTGACATC GAATTCAACT TTAACAGTTT TCATGTTCCG 60
 25 TGATGTTTCa ATAGAATGTG TGTGTTGTAC TTGCGCATT TATTTCAC CTAAATTACT 120
 TAAGTTTCCT GTAATACTAG AAATGTCAGG TCGTTTAAAT GTAGGTTGAA ATGCATCAAC 180
 30 TACTTTATCT GCAACATTAG AACATTACG GATAACTTTA CTTGAATGAT TATCTATACC 240
 TTTAACGAAA CCTAACATTG AATACATACC AACATCCATG AATTCACGTG AAGGTGAGTG 300
 AATACCTAGC GCTCTTTTGG CTGCATTTAA AGCACCTTTT GCTACACTAG CTGCTTTTTC 360
 35 AGCTAAGTCT CTAGCCATAT TACCAATACC TCTCATCAA CCACGGATCA TATCAGCACC 420
 TGCTGATACA AAGTCATCCA CAAAGCTTTT AACTTTATTT ACTGCATTTG TCATACCTTG 480
 ACTAACTTTG TTTACAACAT TAACGAATCC TTGAACAACT CTATTAACAA rGTTAATTAG 540
 40 CGTACTtTGt ATAGTAGATA CCCaTnGCAT ACCTTTAGTG ACmATGAAGT TCCAAGCTTG 600
 AGACATTTTG TCTGATATAG TTGAAACAAC TTGTGTGAAT ATGCTTACAA CTTTATTCCA 660
 AATTGTCGTT AATATACCAG ATAAGAACT CCAAATCGTA TTCCATATAT TAGAAATAAA 720
 45 ACTCCATGCC GCTTGTAACG CAGTAGATAT AGCTGTAGTG ATAGCGTTCC AAACCTTAGT 780
 TGCCACAGTA ACTATAGTGT TCCACAACGT TTGTAAGAAC GTCCAAATAG CGTTCCAAAT 840
 50 TGTTATTGCG ATAGTCATAA TTGTGGTAAA CACTGTAGTT ATTACAGTGA CTAACAAATT 900
 CCAAATCGTA GTAGCGATTG TAATTATCGT ATTCCAGATT GTACTTAAGA ACGTCCAAAT 960
 AGCTGTCCAT ATCGTCATAA CTATTGTCAT TATCGTCGTG AAAACAGTTG TAATGATTGT 1020

	ATAAGCGACT ATTTGATTCC AAACAATCAT TATAAAATTG TAAACATTCC ATACTGCTGT	1140
	AGTGATAGCT GTTAAAATAG CATTCCATAC AACCGAAGCT ACAGCTTTTA ATACATTCCA	1200
5	AACATTAACC ATAAACGTTT TTATCGCATT CCAAGCATTT ATAATAAAGT TTCTGAATCC	1260
	TTCATTTTTA TTCCACAATA AAACGAATAT AGCTATTAAT GCAGCAATTA CACCAATTAC	1320
10	TATTGTTATT GGACCGCCTA AAATACCAAA CACAGTACT AGTCCTGTGA TAGCATTCT	1380
	AATTAATCCA ATCTTACCGA ATAACAATTG GAATATAACT GATATAATTT TTAATGGTCC	1440
	TTTTAATAAC ATGAACGCAC CTTTTAAAT TGTTAATCCC GCTCTTAATA AACCGAACTT	1500
15	ACTTACTAAT GCAATGTTT TACCTATTAA TCCGCCACCC ATAAAGTTAG ATACAGCAAG	1560
	AATAATCGGT ATTAAAAATC TAAATGCACC AACTAAAGTT ATAATGACAC CAACTAATTG	1620
	TGCTGTAGCT GGATGCGCCT CAAACAAGTT AGCTATCCAA CCAGTTATTG CAACTGCAAC	1680
20	GCGTAATACT GCACTAGCTA TAGGAGCCAT TGCTGTTGCG AATGCAmTA ATCCTCTTGC	1740
	GATGTTTCCA ATCAATTGCA TTATTAGTGG TCCATTTGTT TGTATATAAC TGACAAAGTC	1800
25	TTTAAACCCT TGAGATTGTC CTACTTGTTT AGACCATTCC CTAAACTTAG CTGTCATTTG	1860
	TTCAAGAGAT TGGAATATGC CAGTTGATGA TCCGCTGAAT GCATTCATCA AATTGTTAAT	1920
	TCCAACGAAA ACATTTTGA AAATATTACC AATGATAGGT AAGTTTGT TTGTGTATTC	1980
30	AATAAACGA GTTATCGAAT TTTCTCCAGC TGCACTATTA GCCCAGTTAG AGAAAGATTG	2040
	ACCTAATCTA TCCAACCAAT CAGCCGACCA TTGAAACAGT GGTGCTAATT GCGTGAATAC	2100
	ATTGACTAAT CCGTCACCAA AACCACCTGC AGCACTTAAT AGCTTGTTAA ATACCGAAAC	2160
35	ACCCGTTGTA TTCATCATAT TAAAGAATCT TGAAGCTACA CTGCTATTTT CAGCCCATT	2220
	AAGCAGCCTT TGAGACGCTT CTTCCATTCC TCTTGAAATA CCACTAAAAA ACGGTTGTAA	2280
	GCTCTGCATT GCAGTTTTAA CAGTATTTAA ACCATTTGCA AGAGTTGTGA AGATAGCGGA	2340
40	TTGATTTTGC TTTATAATAT CAGTCCATGC TGACTTTACG CCATCTAACG CTTTTTTGTA	2400
	TTGTTTTGTT GCTGAGCTAG CTTGTAAAGT GCCATCATTAGCATCTTTA TAGCGCTGAT	2460
45	AGCCATTGCG CCAAACGCTA CAAATCCTGC TCCCGCTATT GCTACGGCAC CACCTAAAGC	2520
	AAGTACACCA CCAGTTAACA CTTTGATAGC GTTTAATAGC GCAAATACTA CAGGTACTAC	2580
	GCTCGCTATT ACAGGTATTA AGATACTAAA AGATGATGTA AGTAATCCAC CAACCATATT	2640
50	AGAACCTACA GTACCGAACA CACGGAACAT ATTAGCTAAA TTCCCCATCT GTCTTTGAAA	2700
	ATTGTCATTT GCTTTTATTA TGTAGGCATA AGCTTTCTTT AAACCATTAG TATCGACATC	2760
55	TACCTTTGTT GTTTTTTTGT TCGGCAATGC GTCTAATGAT TTTTAAACG CATAAATAGT	2820

	AAGTCTCTCT	TTAGTACGTT	TGATTTTAGA	GTTAGCAACA	CCATTGTCCA	CGTCTATAAT	2940
	AGCTTTGGCT	TTAGACCTAT	TTAATGCTTC	GAGACTAGCT	TTAGATACTT	TTAACACTCG	3000
5	ATTGAATTTA	CTGTTATCTG	CATTGACGTC	AATATTGACA	CGTTTCTTTT	CTAATTCTGA	3060
	TAATTTAGCT	TCTGTTTCAG	CGATATCTTT	AATCAACTTT	TGTTTTTGCA	ACTTAACTTC	3120
10	TGGTGTAAC	TCTTTAGAGT	TTAGTTTGTC	TAGTTCAAAA	TTCGATTCTA	GTACCTTTTG	3180
	TTGTAAATCT	TGTATACTAG	CATCTAATTT	AGCTTTTACA	TTTTTGTTAC	TAAAGGCATC	3240
	TAAAGACTTT	TTAGCAACTT	TGATAGTTTT	TTGTAAATTT	TTATCGTTAG	CGTTTAATTC	3300
15	AACATCTTTA	GTTTGATCTG	CTACTCGTTT	AAATCTTTGC	ACAGACTTAA	CCGCACTATC	3360
	AATTTGCCTT	TTGAATTTGG	CTACACTAGC	TTCAATAGTC	GCTTTAATTT	TATATTCCGT	3420
	CACATTAACA	CCTCTCTTTC	TATTGCTTAT	TAAATTCTGC	TATAACTTTA	AAGAATTCAT	3480
20	TATTTTGTGG	TTCGTATTCA	TCACGTTCCG	TACTAAATCT	TATATCTTTA	CCTTCGTTAA	3540
	GCCGTTGGAT	ATTTTCTTCA	TAAGGCAATA	CGTCGTTTGC	ATTGTTAAAA	ACATATTCCCT	3600
	CTTTAGGTTT	ATTTTCTGTC	CCAACATTTT	TAGTAGCTGC	AGCATCACGA	ATAGCAAACG	3660
25	CAAGTTTGTA	ACGTTCTGAAT	TCTTGGGTGA	GCATTTTATA	CTCTTTTCGCA	TACATTCGAT	3720
	AGTTATATTC	TGTTAATGTC	ATTTGCTCAA	TAACGTTCAA	ATCTGTAATA	CCAAGTGTTG	3780
30	ACATACAAGT	TATAACGATT	CTGTCGTAAG	TTATTAGGcT	TCCGCTGGTT	TTTCTTCCGT	3840
	TTCCACTACT	TCGACTAGGT	TTCCGGTCAT	AGGTCGCTTT	CCCAAcTCCG	TTAAATATATC	3900
	CGAACCGAAT	TCTTCTAGTC	CGATATTTTC	TGCGATTTCA	TCTAATGCTT	CATCAATGTT	3960
35	ATTAATAGTA	ATTGCTTGTT	TTTTTAAGTG	AGATGTAGCT	GCGATTAAAA	CTTCGCCAAT	4020
	CACAACCGGA	TTTCCACTTT	CTAAACCTAC	AGGCAACATT	GATACACCTT	GACCGATAGA	4080
	AGCTTGTTC	ACTTTTAAAC	CTAATCGGTT	ATCGATTTCT	CTTAAAAATT	TAAAACCAAA	4140
40	ACTTAATTCT	AATGACTTTC	CGTTAATTTT	TACATTCATA	ACTTAAAATC	TCCATTCATA	4200
	ATTAATTTAA	ACAAAATAAA	mArGCTTAAC	GCCCTATTTT	TATACCTCTC	TTGGTGCAAC	4260
45	CGGTGGTGAA	TCTACTTTAG	GTTGTGGAAT	TGCTGTTAAA	TCTTCGCCAG	TTAATGCATC	4320
	TGCTTTTGTA	GTGTCGTGGA	ATCTGTATcC	AGTCGCCTTA	AGTTTCTTTG	TTACAGCCTC	4380
	AGGTAGTGTT	GCAAATCCAC	GTTGGAAACG	ACCATTCACT	CCATATTCAT	ATTCATATTC	4440
50	ATCAATACCG	TTAGCTTCTG	CTTTTAATTC	AAATTTATTG	TGGAAACCTT	GGAAATATTT	4500
	CGCTTTAAAT	TTAGCGGAAT	CCCCATTTTT	GCCTGGTATT	CTACTTTCAA	CTTCCCAAGC	4560
55	TTCATACAAT	ACGCGATCTA	CAACTGCATC	TTCAATTTCA	TCTGCAAAAT	CGTCACCATA	4620

GTCCATTGTA TCCTCTGTAT CTGTATCAGC TTCATGTGAT AAGCCGTATT CAGTTAAAAA 4740
 AAGCATTTTA GTAGCATCTA CTTTTTCGCC AGCTTTTCTA AATAAAATAA TACGATCATT 4800
 5 ACTATTTTTC ATATTTGCCA TTCAATATTC CTCCGTTTTT TAAAATGTTT TGTAAGATAT 4860
 CGTTACTGAT GTGTGTAGCA ATTCTTGATT GGTAGTATCA TCAACTAACT GTGTGATGTT 4920
 10 AGTATCTTCT TCTTCAAAGT CATAATCGTT TGTTTTAACG CTAGGTGTTA AATCATCAAT 4980
 ACATCTTTTA ACAAGTCCGT CATGATGTCC TAAATCATCG CTTACACTCC AAATATCAAT 5040
 AACTAAATTC GTATCGCCAG AATAACTATC AAACGTGTAC TTA CTCTCTAT TTGACTCCGG 5100
 15 CATTTTTTATT ACAAAAAAAG GATACGGAAT CTCTTGTTGC ATCTCTTTAC GAGAAATAAC 5160
 AGGGAATCCA TATCCTTGTA GCGTTTCATA CGCTTTTATTA TAAAGTTGTA AGTTCGGTGT 5220
 CATGCTTTTA TCTCCTATTC AAACAACGCT TTCAATCTTT CTACAGTTGA TTTCCTAATC 5280
 20 ACTTCGTATA CCGGCCACAT AAAAGGTTCA GCCTCCATGT ATCGAGTACC AAATTCTAAG 5340
 AAACCACTAT AAGCTGCGTG CGATGTGATA GTGTATTGCA AATCGCCAGT TTTTTTATAT 5400
 CTGATATTGC GTGATAAATT ACC 5423
 25

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6251 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAACGCAGAT GTTCAATTAG AACCAGTCTA TCGTATTAAG GAAGGTATTA AACAAAAGCA 60
 AATAAGAGAC CAAATTAGAC AAGCGTTAAA TGATGTGACA ATTCATGAAT GGTAACTGA 120
 40 TGAAC TAAGA GAAAAATATA AATTAGAGAC CTTGGACTTT ACTTTGAACA CATTACATCA 180
 TCCTAAAAGT AAAGAGGATT TATTACGTGC TCGTAGAACC TATGCATTTA CTGAACTGTT 240
 45 TTTATTCGAA TTACGTATGC AATGGCTAAA TAGATTAGAA AAGTCATCTG ACGAAGCAAT 300
 TGAAATTGAT TATGACATAG ACCAAGTTAA ATCATTATT GATCGTTTAC CTTTGAAC 360
 AACTGAAGCA CAGAAATCCA GTGTTAATGA AATTTT TAGA GATT TAAAAG CACCAATACG 420
 50 TATGCATCGA TTA CTTC AAG GTGATGTAGG TTCAGGAAAA ACAGTAGTTG CTGCAATTTG 480
 TATGTATGCG TTA AAAACTG CTGGTTATCA ATCAGCATTG ATGGTACCAA CTGAAATTTT 540
 AGCAGAGCAA CATGCTGAAA GTTTAATGGC TTTATT TGGA GATTCTATGA ACGTTGCATT 600

55

	TACGATTGAT	TGTTTAATTG	GAACCCATGC	TTTGATTCAA	GATGATGTGA	TTTTCCATAA	720
	TGTTGGTTTA	GTAATTACAG	ATGAACAACA	TCGATTGGGT	GTGAATCAAC	GCCAGCTTTT	780
5	AAGAGAAAAA	GGTGCAATGA	CGAATGTGTT	ATTTATGACA	GCAACGCCGA	TACCAAGAAC	840
	ACTAGCAATA	TCAGTTTTTG	GTGAGATGGA	TGTGTCTTCA	ATTAAACAAT	TACCAAAAAGG	900
10	TCGTAAACCT	ATCATTACTA	CTTGGGCCAA	GCATGAGCAA	TACGATAAAG	TTTTGATGCA	960
	AATGACCTCA	GAGTTGAAAA	AAGGTCGTCA	AGCATATGTC	ATTTGCCCGC	TAATAGAAAG	1020
	TTCTGAGCAT	CTCGAAGATG	TTCAAAATGT	TGTCGCATTG	TACGAGTCTT	TACAACAGTA	1080
15	TTATGGTGTT	TCCCGTGTAG	GGTTATTGCA	TGGTAAGTTA	TCTGCCGATG	AAAAAGATGA	1140
	GGTCATGCAA	AAGTTTAGTA	ATCATGAGAT	AAATGTTTTA	GTTTCTACTA	CTGTTGTTGA	1200
	AGTAGGTGTT	AATGTACCGA	ATGCAACTTT	TATGATGATT	TATGATGCGG	ATCGCTTTGG	1260
20	ATTATCAACT	TTACATCAGT	TACGCGGTCT	TGTAGGTAGA	AGTGACCAGC	AAAGTTACTG	1320
	TGTTTTAATT	GCATCCCCTA	AAACAGAAAC	AGGAATTGAA	AGAATGACAA	TTATGACACA	1380
	AACAACGGAT	GGATTTGAAT	TGAGTGAACG	AGACTTAGAA	ATGCGTGGTC	CTGGAGATTT	1440
25	CTTTGGTGTT	AAACAAAGTG	GaTTGCCAGA	TTTCTTAGTT	GCCAATTTAG	TTGAAGATTA	1500
	TCGTATGTTA	GAAGTTGCTC	GTGATGAAGC	AGCTGAACTT	ATTCAATCTG	GCGTATTCTT	1560
30	TGAAAATACG	TATCAACATT	TACGTCATTT	TGTTGAAGAA	AATTTATTAC	ATCGTAGTTT	1620
	TGACTAATTG	CCATGCTGAT	TTGTCAATTT	GAGTGCAACa	CTTCGTTAAT	TGAGTGATAT	1680
	GACACTTGAA	CTATTTAAAT	GTAAAGTGGT	ATTTTAACAA	TTTATAAATT	TTCGACTAAA	1740
35	TAATAGCTAA	ATATTACAGT	TATTTGTTGA	GTCGGTTAAA	TAGAAAGTGT	TATGATATGT	1800
	GAGGAATGTT	TAAGACTAGG	TACTAAAAAA	TGAGGGGTGA	GACGTTGAAA	CTAAAGAAAAG	1860
	ATAAACGTAG	AGAAGCAATC	AGACAACAAA	TTGATAGCAA	TCCCTTCATC	ACAGACCATG	1920
40	AACTAAGCGA	CTTATTTCAA	GTGAGTATAC	AAACAATTCT	TTtAGaTCGC	ACTTATTTAA	1980
	ACATACCAGA	ATTAAGGAAG	CGTATTAAAT	TAGTTGCTGA	AAAGAATTAT	GACCAAATAA	2040
45	GTTCTATTGA	AGAACAAGAA	TTTATTGGTG	ATTTGATTCA	AGTCAATCCa	AATGTTAAAG	2100
	CGCAATCAAT	TTTAGATATT	ACATCGGATT	CTGTTTTTCA	TAAAACTGGA	ATTGCGCGTG	2160
	GTCATGTGCT	GTTTGCTCAG	GCAAATTCGT	TATGTGTTGC	GCTAATTAAG	CAACCAACAG	2220
50	TTTTAACTCA	TGAGAGTAGC	ATTCAATTTA	TTGAAAAAGT	AAAATTAAAT	GATACGGTAA	2280
	GAGCAGAAGC	ACGAGTTGTA	AATCAAACCT	CAAAACATTA	TTACGTCGAA	GTAAAGTCAT	2340
	ATGTTAAACA	TACATTAGTT	TTCAAAGGAA	ATTTTAAAAT	GTTTTATGAT	AAGCGAGGAT	2400
55							

	TTAGAAGCCG	TACAAAAGGC	TGTTGAAGAC	TTTAAAGATC	TAGAAATTAT	ACTTTTCGGT	2520
	GACGAAAAAA	AGTATAATCT	GAACCATGAA	CGAATCGAAT	TTAGACATTG	TTCTGAAAAG	2580
5	ATTGAAATGG	AAGATGAGCC	TGTTAGAGCG	ATTAAACGTA	AAAAAGATAG	CTCAATGGTA	2640
	AAAATGGCTG	AAGCTGTGAA	ATCTGGTGAA	GCAGATGGAT	GTGTGTCAGC	AGGTAATACT	2700
10	GGTGCTTTAA	TGTCAGCTGG	TTTATTCATT	GTTGGACGTA	TTAAAGGTGT	AGCTAGACCG	2760
	GCTTTAGTAG	TAACATTGCC	AACGATTGAT	GGAAAAGGTT	TTGTCTTTTT	AGACGTTGGT	2820
	GCAAATGCTG	ATGCTAAACC	TGAACACTTA	TTACAGTATG	CGCAACTAGG	GGATATTTAT	2880
15	GCTCAAAAAA	TTAGAGGTAT	TGATAATCCG	AAAATCTCAT	TATTAAATAT	AGGAACCGAG	2940
	CCAGCTAAAG	GTAATAGTTT	AACGAAAAAA	TCATATGAGT	TATTAAATCA	TGATCATTCA	3000
	TTGAATTTTG	TTGGGAATAT	TGAAGCGAAG	ACATTAATGG	ATGGCGATAC	AGATGTTGTA	3060
20	GTTACCGATG	GCTATACTGG	GAACATGGTC	CTTAAAAATT	TAGAAGGTAC	TGCAAAATCA	3120
	ATCGGTAAAA	TGTTAAAAGA	TACGATTATG	AGTAGTACTA	AAAATAAATT	AGCAGGTGCA	3180
25	ATATTGAAGA	AAGATTTAGC	TGAATTCGCT	AAAAAGATGG	ATTACTCAGA	ATACGGTGGT	3240
	TCCGTATTAT	TAGGATTGGA	AGGTACTGTA	GTTAAAGCAC	ACGGTAGTTC	AAATGCTAAA	3300
	GCTTTTTATT	CTGCAATTAG	ACAAGCGAAA	ATCGCAGGAG	AACAAAATAT	TGTACAAACA	3360
30	ATGAAAGAGA	CTGTAGGTGA	AtCAAATGaG	TaAAACAGCA	ATTATTTTTTC	CGGGACAAGG	3420
	TGCCCCAAAA	GTTGGTATGG	CGCAAGATTT	GTTTAACAAC	AATGATCAAG	CAACTGAAAT	3480
	TTTAACTTCA	GCAGCGAACA	CATTAGACTT	TGATATTTTA	GAGACAATGT	TTACTGATGA	3540
35	AGAAGGTAAA	TTGGGTGAAA	CTGAAAACAC	ACAACCAGCT	TTaTTGaCGC	aTAGTTCGGC	3600
	ATTATTAGCA	GCGCTAAAAA	ATTTGAATCC	TGATTTTACT	ATGGGGCATA	GTTTAGGTGA	3660
	ATATFCAAGT	TTAGTTGCAG	CTGACGTATT	ATCATTTGAA	GATGCAGTTA	AAATTGTTAG	3720
40	AAAACGTGGT	CAATTAATGG	CGCAAGCATT	TCCTACTGGT	GTAGGAAGCA	TGGCTGCAGT	3780
	ATTGGGATTA	GATTTTGATA	AAGTCGATGA	AATTTGTAAG	TCATTATCAT	CTGATGACAA	3840
45	AATAATTGAA	CCAGCAAACA	TTAATTGCCC	AGGTCAAATT	GTTGTTTCAG	GTCACAAAGC	3900
	TTTAATTGAT	GAGCTAGTAG	AAAAAGGTAA	ATCATTAGGT	GCAAAACGTG	TCATGCCTTT	3960
	AGCAGTATCT	GGACCATTCC	ATTCATCGCT	AATGAAAGTG	ATTGAAGAAG	ATTTTTC AAG	4020
50	TTACATTAAT	CAATTTGAAT	GGCGTGATGC	TAAGTTTCCT	GTAGTTCAAA	ATGTAAATGC	4080
	GCAAGGTGAA	ACTGACAAAG	AAGTAATTAA	ATCTAATATG	GTCAAGCAAT	TATATTCACC	4140
55	AGTACAATTC	ATTAAC TCA	CAGAATGGCT	AATAGACCAA	GGTGTTGATC	ATTTTATTGA	4200

	AACATCAATT CAAACTTTAG AAGATGTGAA AGGATGGAAT GAAAATGACT AAGAGTGCTT	4320
	TAGTAACAGG TGCATCAAGA GGAATTGGAC GTAGTATTGC GTTACAATTA GCAGAAGAAG	4380
5	GATATAATGT AGCAGTAAAC TATGCAGGCA GCAAAGAGAA AGCTGAAGcA GTAGTCGAAG	4440
	AAATCAAAGC TAAAGGTGTT GACAGTTTTG CGATTCAAGC AAATGTTGCC GATGCTGATG	4500
10	AAGTTAAAGC AATGATTAAA GAAGTAGTTA GCCAATTTGG TTCTTTAGAT GTTTTAGTAA	4560
	ATAATGCAGG TATTACTCGC GATAATTTAT TAATGCGTAT GAAAGAACAA GAGTGGGATG	4620
	ATGTTATTGA CACAACTTA AAAGGTGTAT TTAAGTGTAT CCAAAAAGCA ACACCACAAA	4680
15	TGTTAAGACA ACGTAGTGGT GCTATCATCA ATTTATCAAG TGTGTTGGA GCAGTAGGTA	4740
	ATCCGGGACA AGCAAATAT GTTGCAACAA AAGCAGGTGT TATTGGTTTA ACTAAATCTG	4800
	CGGCGCGTGA ATTAGCATCT CGTGGTATCA CTGTAAATGC AGTTGCACCT GGTTTTTATTG	4860
20	TTTCTGATAT GACAGATGCT TTAAGTGTAT AGCTTAAAGA ACAAATGTTG ACTCAAATTC	4920
	CGTTAGCACG TTTTGGTCAA GACACAGATA TTGCTAATAC AGTAGCGTTC TTAGCATCAG	4980
	ACAAAGCAAA ATATATTACA GGTCAAACAA TCCATGTAAA TGGTGAATG TACATGTAAT	5040
25	ATATTTGAGC TAAAGCTCAT TGACGCAGTG GTTGACTGGT CATCCAATGG AGAATTGTCT	5100
	GACCTAGTCA ACTTTGCGGG GGAAATTCTA AGCAACCTAG ATAAGGTTCC AGAATTTCTC	5160
30	CCTAAGAAAC ACTAATCAAT aAATTGWIAA GTGTTTCTAA AATTTCTACT TGTTTTTTAG	5220
	AATTTAAAAT GGGAAAATAT AGTAGTCTAT GTATAGGCAT TTTTAAAGGA GGTGAATCGA	5280
	CGTGGAATAT TTCGATAAAG TAAAAGATAT CATCGTTGAC CgTTTAGGTG TAGACGCTGA	5340
35	TAAAGTAACT GAAGATGCAT CTTTCAAAGA TGATTTAGGC GCTGACTCAC TTGATATCGC	5400
	TGAATTAGTA ATGGAATTAG AAGACGAGTT TGGTACTGAA ATTCCTGATG AAGAnGCTGA	5460
	AAAAATCAAC ACTGTTGGTG ATGCTGTTAA ATTTATTAAAC AGTCTTGAAA AATAATAAAT	5520
40	CTTACATCTG GGTCGTCAGT ATTGTCGACT CAGTTTTTTT CTTTAATTAT CAATAGTTTT	5580
	AACGTAAAAT TAAAGATGAT TCAAGAGCAA CACATAAAGG AGATAAAATA ATGTCTAAAC	5640
	AAAAGAAAAG TGAGATAGTT AATCGTTTTA GAAAGCGCTT TGATACTAAA ATGACAGAGT	5700
45	TAGGCTTTAC TTATCAAAAT ATTGATTTAT ACCAACAAGC ATTTTCGCAT TCGAGTTTTA	5760
	TTAATGATTT TAATATGAAT CGTTTAGACC ATAATGAGCG TTTAGAGTTT TTGGGTGATG	5820
50	CGGTATTAGA ATTGACGGTT TCACGATATT TATTTGATAa ACATCCCAAC TTGCCAGAAG	5880
	GGAATTTAAC AAAAATGCGT GCCaCTATTG TATGTGAGCC CtCACTkGTA ATATTTGCGA	5940
	ATAAAATTGG ATTGAACGAA ATGATTTTAC TTGGTAAAGG TGAAGAGAAA ACAGGGGGAC	6000

55

ATCAAGGACT AGATATAGTT TGGAAATTTG CTGAGAAAGT CATTTCCTCA CATGTAGAAC 6120
 AAAATGAGTT ATTAGGCGTG GTAGATTTTA AAACACAATT CCAAGAATAT GTGCACCAGC 6180
 5 AAAATAAAGG TGATGTAACC TATAATTTAA TAAAAGAAGA GGGACCGGCA CATCATCGTC 6240
 TATTCACTTC A 6251

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4920 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

20 ACCTACTGAA GTTGCTAATT TTTTGAGCA ACTAAGCACT GAAATTGAAC GTCTTAAAGA 60
 AGATAAAAAA CAACTTGAAA AAGTAATCGA AGAGAGAGAT ACTAATATTA AGTCTTATCA 120
 AGACGTGgCA TCAATCTGTA AGTGaTGCTT TGATACAAGC TCAAAAAGCT GGTGAAGAAA 180
 25 CTAAGCAAGC TGCAGAGAAA CAAGCTGAAG CGATTATAGC TAAGGCAGAA GCGCAAgcTA 240
 ATcAAATGGT TGGTGACGCG GTAGAAAAAG CACGCCGTTT AGCATTCCAG ACTGAAGATA 300
 TGAAACGTCA ATCAAAAGTA TTTAGATCGC GTTTCCTAT GTTAGTTGAA GCGCAATTAG 360
 ACTTATTAAA AAACGAAGAT TGGGATTACT TGTTGAATTA TGATTTAGAC GCTGAACAAG 420
 TGACGCTTGA AAATATTCAT CATTGTCATG AAAATGATTT AAAGCCAGAT GAAGTTGCAG 480
 35 CAAATGCACA AAATAATGCA TCAAATACAC CAGACAATAA TCAACAATCC AATGATTCAG 540
 AAACAACATA GAAGTAAGAA TTAAATAAAG ACAGACGCGT AATATACATT TAACTTTTCa 600
 CAGCGAATTA GGTAATGGTG AGAGCCTAGT AAAAGCATGT ATGTTATATC ACTGGCTTTT 660
 40 TAATATTTAA ATAATGTAAT GAGAGAACTC TAAGTTGAGT TAATAAGGGT GGTACCGCGA 720
 GCAATCGTCC CTTTTAATTT AACTTAGAGT TTTTAAATT TTTAAGGAGT GAAAAAATG 780
 GATTACAAAG AAACGTTATT AATGCCTAAA ACAGATTTCC CAATGCGAGG TGGTTTACCA 840
 45 AACAAGGAAC CGCAAATTCA AGAAAAATGG GATGCAGAAG ATCAATACCA TAAAGCGTTA 900
 GAAAAAATA AAGGTAACGA AACATTCATT TTACATGATG GCCCACCATA CGCGAATGGT 960
 50 AACTTACATA TGGGACATGC CTTGAACAAA ATTTTAAAAG ACTTTATTGT ACGTTATAAA 1020
 ACTATGCAAG GGTTCTATGC ACCATACGTA CCAGGTTGGG ATACACATGG TTTACCAATT 1080
 GAACAAGCAT TAACGAAAAA AGGTGTTGAC CGAAAGAAAA TGTCACAGC TGAATTCCGT 1140

	TTAGGTGTTT	GTGGTGACTT	TAATGATCCA	TATATTACAT	TAAAACCTGA	ATACGAAGCT	1260
	GCACAAATTC	GTATTTTGG	AGAAATGGCA	GATAAAGGTT	TAATTTATAA	AGGTAAAAAG	1320
5	CCAGTTTATT	GGTCTCCTTC	AAGTGAGTCT	TCATTAGCAG	AAGCAGAAAT	TGAATATCAC	1380
	GATAAACGTT	CAGCATCAAT	TTACGTTGCA	TTTGACGTTA	AAGATGACAA	AGGTGTCGTT	1440
10	GATGCAGATG	CTAAATTTAT	TATCTGGACA	ACAACGCCAT	GGACAATTCC	ATCAAATGTT	1500
	GCGATTACCG	TTCATCCTGA	ATTAAATAT	GGTCAATACA	ATGTAAATGG	cGAAAAATAT	1560
	ATTATTGCAG	AAGCCTTGTC	TGACGCTGTA	GCAGAAGCAC	TGGaTTGGGA	TAAAGCATCA	1620
15	ATCAAATTAG	AAAAAGAATA	CACAGGTAAA	GAATTAGAGT	ATGTTGTAGC	ACAACATCCA	1680
	TTCTTAGACA	GAGAATCGTT	AGTGATTAAT	GGTGATCATG	TTACTACAGA	TGCTGGTACA	1740
	GgTTGTGTAC	ATACAGCACC	AGGTCACGGG	GAAGATGACT	ATATTGTTGG	TCAAAAATAT	1800
20	GAATTGCCAG	TAATTAGTCC	AATCGATGAT	AAAGGTGTAT	TTACTGAAGA	AGGCGGCCAA	1860
	TTTGAAGGGA	TGTTCTATGA	TAAAGCTAAT	AAAGCCGTTA	CTGATTTATT	AACAGAAAAA	1920
	GGTGCACTAT	TAAAATTAGA	CTTTATTACA	CATAGCTATC	CACACGACTG	GAGAACAAAA	1980
25	AAACCTGTAA	TCTTCCGTGC	TACACCACAA	TGGTTTGCCT	CAATCAGTAA	AGTAAGACAA	2040
	GATATTTTAG	ATGCAATCGA	AAATACAAAC	TTCAAAGTAA	ATTGGGGTAA	AACACGTATT	2100
30	TACAATATGG	TTCGTGACCG	TGGCGAATGG	GTTATTTCTC	GTCAACGTGT	GTGGGGTGTA	2160
	CCGTTACCAG	TATTTTATGC	TGAAAATGGC	GAAATTATCA	TGACGAAAGA	AACAGTGAAT	2220
	CATGTTGCTG	ATTTATTTGC	AGAACACGGT	TCAAATATTT	GGTTTGAAAG	AGAAGCGAAA	2280
35	GACTTACTAC	CAGAAGGATT	TACACATCCA	GGCAGCCCTA	ACGGTACATT	TACTAAAGAA	2340
	ACAGACATTA	TGGACGTTTG	GTTTGATTCT	GGTTCATCAC	ACCGTGGCGT	GTTGGAAACA	2400
	AGACCGGAAT	TAAGTTTCCC	AGCGGATATG	TATTTAGAAG	GTAGTGACCA	ATATCGTGGT	2460
40	TGGTTCAACT	CTTCTATCAC	AACTTCAGTT	GCTACAAGAG	GAGTATCACC	TTATAAATTC	2520
	TTACTTTCTC	ATGGTTTTGT	TATGGACGGT	GAAGGTAAGA	AAATGAGTAA	ATCTTTAGGT	2580
45	AATGTGATTG	TACCTGACCA	AGTGGTTAAA	CAAAAAGGTG	CTGATATTGC	GAGACTTTGG	2640
	GTAAGTAGTA	CGGACTATTT	AGCTGATGTT	AGAATTTCTG	ATGAAATTTT	AAAACAAACA	2700
	TCTGATGTTT	ATCGTAAAT	CAGAAATACA	TTAAGATTTA	TGTTAGGTAA	CATTAAACGAT	2760
50	TTCAATCCTG	ACACAGATAG	CATTCCTGAA	TCAGAGTTAT	TAGAAGTGGA	TCGTTACTTG	2820
	CTAAATCGTT	TACGTGAATT	TACTGCAAGT	ACGATTAACA	ACTATGAAAA	CTTTGACTAC	2880
	TTAAATATTT	ATCAAGAAGT	TCAAACTTT	ATCAATGTTG	AGTTAAGTAA	TTTCTATTTG	2940

55

	CAAACAGTGT TATATCAAAT TTTAGTTGAT ATGACGAAGT TGTTAGCACC AATCTTAGTG	3060
5	CATACAGCTG AAGAAGTTTG GTCTCATACA CCACATGTTA AAGAAGAAAAG TGTTCACTTA	3120
	GCAGACATGC CTAAAGTTGT AGAAGTAGAT CAAGCTTTAT TGGATAAATG GCGTACATTT	3180
	ATGAATTTAC GTGATGATGT GAACCGTGCA TTAGAAACTG CTCGTAATGA AAAAGTTATT	3240
10	GGTAAATCAT TAGAAGCTAA AGTTACGATT GCTAGTAACG ATAAATTTAA TGCATCTGAA	3300
	TTCTTAACTT CATTTGATGC ATTACATCAA TTATTTATCG TGTCAACAAGT TAAAGTTGTA	3360
	GATAAGTTAG ACGATCAGGC AACAGCTTAT GAACATGGTG ATATTGTCAT CGAACATGCA	3420
15	GATGGTGAAA AATGTGAAAG ATGTTGGAAC TATTCAGAGG ATCTTGGTGC TGTTGATGAA	3480
	TTGACGCATC TATGTCCACG ATGCCAACAA GTTGTAATAA CACTTGTATA ATTGAAATTG	3540
	TATAAAGTAC TCATACAGAT GATATAAATT AAAGCTCTCT TCATAATCAT GTTGTAGTTT	3600
20	TTGTTGACAT GATGAAGAGA GTTTTTTTGT GAATAAAAAA ATGACCAAGT TACCGGTCAT	3660
	ATATGTAAAA AATGTGCGAT TTAATAAAT AAAAATTATT CAGGAATGGT ACAAATTCTC	3720
25	TGAGGCATAT AAATGCGTTA TAGTTGCTAT TCTCAATTAT GTTCGCGATA ATTTTAAGTA	3780
	AAAGTAAGCA CAGATATTGA ATTTGATAGG AGTTAATTGA ATGTATCATA ACAGTAACGC	3840
	AAACTTTGTC AATGGTATCA CTTTAAATGT GAGAGATAAG AATGAATTAA AGCCATTTTA	3900
30	TGAGGACATA TTAGGATTAA ATATTATAAA TGAGACATTA ACATCGATAC AATATGAAGT	3960
	AGGTCAAAAT AATCATGTCA TTACACTTGT TGAATTACAA AATGGACGTG AACCTTTAAT	4020
	GTCCGAAGCG GGAAGTTTC ATATCGCAAT TAACTACCT CAAATTAGTG ATTTAGCTAA	4080
35	TTTACTAATT CATTTAAGCG AATATGATAT TCCAGTTAAC GGAGGTATAC AGCCTGCTTC	4140
	GTTATCATT TTTTTTGAAG ACCCGGAAGG AAACGGTTTT AAATTTTATG TTGATAAAGA	4200
	CGAAGCGCAA TGGACGAGGC AAAATAATTT AGTAAAAATT GATATTAGAC CATTAAATGT	4260
40	ACCGAGATTA GTGAGTCATG CAACAAAATT GTTATGGTTA GGTATTCCAG ATGACGCTAT	4320
	TATAGGTGCA TTGCATATTA AGACAATTCA TTTATCAGAG GTAAAAGAGT ACTACCTCGA	4380
45	TTATTTTGGA TTAGAGCAAT CGGCATATAT GGATGATTAT TCAATATTTT TAGCATCGAA	4440
	TGGCTATTAT CAACATTTGG CCATGAATGA TTGGGTATCA GCAACGAAAC GTGTAGAAAA	4500
	TTTTGATACG TATGGATTAG CAATTGTTGA CTTTCATTAT CCTGAAACAA CACATTTAAA	4560
50	TTTACAAGGT CCGGATGGTA TCTATTATCG CTTTAATCAT ATCGAAGTTG AAGATTAGTA	4620
	TATACTTTGA ATGGACGAAC CATATAATGA ATCGTTTTTA ATGATCTTTT TATACAAGTT	4680
55	ATGAAGGAGG CTGGGACATT AAGTTCTTAG GCAATGTAAA AAGCTGATTT CTATTAATTA	4740

TTTTCCTTAT ATTAATTGCC ATTAATACAA AACCTAGCTC TCGTTTAACT TTATTTATTC 4860

CTCGAACTGA CATTTCGnGTG AACTCAAAAT nGCCTACTTn CTTAAATTAC CAATATCTAT 4920

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 626 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TGGATTGCCA TTACATGGAC AAGATTTAAC TGAATCAATT ACACCATATG AAGGTGGTAT 60

CGCTTTTGCA AGTAAACCAT TAATTGATGC TGATTTTATT GGTAAATCTG TATTAAAAGA 120

TCAAAAAGAA AATGGTGCAC CAAGAAGAAC AGTGGGATTA GAATTACTTG AAAAAGGAAT 180

TGCAAGAACT GGTTATGAAG TTATGGATTT AGATGGAAAT ATTATTGGAG AAGTAACTTC 240

AGGAACACAG TCTCCATCAT CAGGAAAATC AATTGCACTT GCAATGATAA AAAGAGATGA 300

GTTTGAAATG GGTAGAGAGT TGCTTGTTCA AGTTCGTAAG CGTCAATTAA AAGCGAAAAT 360

TGTTAAGAAA AATCAAATTG ATAAATAATT AAAAAGGGGT GTGCATTGTG AGTCATCGTT 420

ATATACCTTT AACTGAAAAA GACAAGCAAG AAATGTTACA AACAATTGGT GCAAAATCTA 480

TAGGAGAATT ATTCGGTGAT GTACCAAGTG ACATTTTATT AAATAGAGAT TTAAATATTG 540

CTGAAGGCGA ACGGAGAACA ACGTACTTGA GAAGATTnAA TCGCATTGCA AGCAAGAGTA 600

TCCTAGAGG AACGCGTACA TCGTTT 626

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1126 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

nGGAAGTGGT GTATATATTT GTAATGAGTG TATTGAATTA TGCTCAGAAA TCGTCGAAGA 60

AGAATTAGCT CAAAACACTT CTGAAGCGAT GACAGAATTA CCTACTCCTA AAGAAATTAT 120

GGATCATTTA AACGAATATG TTATTGGTCA AGAAAAAGCT AAAAAATCTT TAGCTGTAGC 180

TGTTTATAAC CACTATAAGC GTATTCAACA ATTAGGACCA AAAGAAGATG ATGTTGAATT 240

AACCTTAGCC AAGACGTTGA ATGTACCATT TGCAATTGCA GATGCGACAA GTTTAACTGA 360
 AGCTGGTTAT GTAGGCGATG ATGTTGAAAA TATCTTGTTG AGATTAATTC AAGCAGCTGA 420
 5 CTTTGACATT GATAAAGCCG AAAAAGGTAT TATTTATGTA GATGAAATTG ATAAAATTGC 480
 ACGTAAATCT GAAAACACAT CTATAACACG TGACGTTTCA GGTGAAGGTG TTCAACAAGC 540
 10 ATTGCTTAAA ATCTTAGAAG GTACGACTGC AAGTGTTCCG CCACAAGGTG GACGCAAACA 600
 TCCAAACCAA GAAATGATTG AAATTGATAC AACAAATATC TTATTTATTC TTGGTGGTGC 660
 CTTTGATGGT ATTGAAGAAG TGATTAAGCG CCGTCTTGGT GAAAAAGTTA TTGGTTTCTC 720
 15 AAGCAATGAA GCTGATAAAT ATGACGAACA AGCATTATTA GCACAAATTC GCCCAGAAGA 780
 TTTGCAAGCC TATGGTTTGA TTCCTGAATT TATCGGACGT GTGCCAATTG TAGCTAATTT 840
 AGAAACATTA GATGTAAC TGTTGAAAAA CATCTTAACG CAACCTAAAA ATGCACTTGT 900
 20 GAAACAATAT ACTAAAATGC TGGAATTAGA TGATGTGGAT TTAGAGTTCA CTGAAGAAGC 960
 TTTATCAGCA ATTAGTGAAA AAGCAATTGA AAGAAAAACA GGTGCGCGTG GTTTACGTTT 1020
 AATCATAGAA GAATCGTTAA TCGATATTAT GTTTGATGTG CCTTCTAACG AAAATGTAAC 1080
 25 GAaGGTAGTT ATTACAGCAC AAACmATTAA TGrAGaACTG AACCAG 1126

(2) INFORMATION FOR SEQ ID NO: 29:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4392 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATTGACTTCT TAGCAATnAA TaTGAGTGAA GAACGTA CTG TTGAAGTACC AGTTCAATTA 60
 40 GTTGGTGAAG CAGTAGGCGC TAAAGAAGGC GCGTAGTTG AACCAACCATT ATTCAACTTA 120
 GAAGTAACTG CTA CTCCAGA CAATATTCCA GAAGCAATCG AAGTAGACAT TACTGAATTA 180
 45 AACATTAACG ACAGCTTAAC TGTTGCTGAT GTTAAAGTAA CTGGCGACTT CAAAATCGAA 240
 AACGATTCAG CTGAATCAGT AGTAACAGTA GTTGCTCCAA CTGAAGAACC AACTGAAGAA 300
 GAAATCGAAG CTATGGAAGG CGAACAACAA ACTGAAGAAC CAGAAGTTGT TGGCGAAAGC 360
 50 AAAGAAGACG AAGAAAAAAC TGAAGAGTAA TTTTAATCTG TTACATTAAA GTTTTTATAC 420
 TTTGTTTAAC AAGCACTGTG CTTATTTTAA TATAAGCATG GTGCTTTTGT TGTTATTATA 480
 AAGCTTAATT AAACCTTATT ACTTTGTACT AAAGTTTAAT TAATTTTAGT GAGTAAAAGA 540

55

	CTTACTAAGC TAAAGAATAA TGATAATTGA TGGCAATGGC GGAAAATGGA TGTTGTCATT	660
	ATAATAATAA ATGAAACAAT TATGTTGGAG GTAAACACGC ATGAAATGTA TTGTAGGTCT	720
5	AGGTAATATA GGTAAACGTT TTGAACTTAC AAGACATAAT ATCGGCTTTG AAGTCGTTGA	780
	TTATATTTTA GAGAAAAATA ATTTTTCATT AGATAAACAA AAGTTTAAAG GTGCATATAC	840
10	AATTGAACGA ATGAACGGCG ATAAAGTGTT ATTTATCGAA CCAATGACAA TGATGAATTT	900
	GTCAGGTGAA GCaGTTGCAC CGATTATGGA TTATTACAAT GTTAATCCAG AAGATTTAAT	960
	TGTCTTATAT GATGATTAG ATTTAGAACA AGGACAAGTT CGCTTAAGAC AAAAAGGAAG	1020
15	TGCGGGCGGT CACAATGGTA TGAAATCAAT TATTAATG CTTGGTACAG ACCAATTTAA	1080
	ACGTATTCGT ATTGGTGTGG GAAGACCAAC GAATGGTATG ACGGTACCTG ATTATGTTTT	1140
	ACAACGCTTT TCAAATGATG AAATGGTAAC GATGGAAAAA GTTATCGAAC ACGCAGCACG	1200
20	CGCAATTGAA AAGTTTGTG AAACATCACG ATTTGACCAT GTTATGAATG AATTTAATGG	1260
	TGAAGTGAAG TAATGACAAT ATTGACAACG CTTATAAAG AAGATAATCA TTTTCAAGAC	1320
25	CTTAATCAGG TATTTGGACA AGCAAACACA CTAGTAACTG GTCTTTCCCC GTCAGCTAAA	1380
	GTGACGATGA TTGCTGAAAA ATATGCACAA AGTAATCAAC AGTTATTATT AATTACCAAT	1440
	AATTTATACC AAGCAGATAA ATTAGAAACA GATTTACTTC AATTTATAGA TGCTGAAGAA	1500
30	TTGTATAAGT ATCCTGTGCA AGATATTATG ACCGAAGAGT TTTCAACACA AAGCCCTCAA	1560
	CTGATGAGTG AACGTATTAG AACTTTAACT GCGTTAGCTC AAGGTAAGAA AGGGTTATTT	1620
	ATCGTTCCTT TAAATGGTTT GAAAAAGTGG TTAACCTCTG TTGAAATGTG GCAAAATCAC	1680
35	CAAATGACAT TCGGTGTTGG TGAGGATATC GATGTGGACC AATTTCTTAA CAAATTAGTT	1740
	AATATGGGGT ACAAACGGGA ATCCGTGGTA TCGCATATTG GTGAATTCTC ATTGCGAGGA	1800
	GGTATTATCG ATATCTTTCC GCTAATTGGG GAACCAATCA GAATTGAGCT ATTTGATACC	1860
40	GAAATTGATT CTATTCGGGA TTTTGATGTT GAAACGCAGC GTTCCAAAGA TAATGTTGAA	1920
	GAAGTCGATA TCACAACCTGC AAGTGATTAT ATCATTACTG AAGAAGTGAT CAGCCATCTT	1980
45	AAAGAAGAGT TAAAACTGC ATATGAAAAT ACAAGACCCA AAATAGATAA ATCAGTGCGC	2040
	AATGATTGA AAGAAACGTA TGAAAGCTTT AAATTATTCG AAAGTACATA CTTTGATCAT	2100
	CAAATACTAC GTCGCTTAGT AGCGTTTATG TATGAAACAC CTTGACAAT TATTGAGTAT	2160
50	TTCCAAAAAG ATGCAATCAT TGCAGTTGAT GAATTTAATC GTATTAAAGA AACTGAAGAA	2220
	AGTTTAACAG TAGAGTCTGA TTCGTTTATT AGCAATATTA TTGAAAGTGG TAATGGATTT	2280
55	ATAGGACAAA GTTTTATAAA ATATGATGAT TTTGAAACAT TGATTGAAGG CTATCCTGTC	2340

	TCATGTAAAC CTGTCCAACA ATTTTATGGG CAATATGACA TTATGCGTTC TGAATTTCAA	2460
	CGATATGTTA ATCAAACTA TCATATCGTG GTTTTGGTCG AAACCGAAAC TAAAGTTGAA	2520
5	CGTATGCAAG CGATGTTAAG TGAAAtGCAT ATTCCATCAA TAACAAAATT GCATCGCTCA	2580
	ATGTCATCGG GGCAAGCAGT GATTATTGAA GGCAGTTTAT CTGAAGGATT TGAACCTACCT	2640
10	GATATGGGAT TAGTTGTCAT TACTGAGCGT GAgcTTTTTA AATCAAAACA GAAAAAGCAA	2700
	CGAAAACGTA CGAAAGCTAT CTCAAATGCT GAAAAAATTA AGTCTTACCA AGATTTAAAT	2760
	GTGGGAGATT ATATTGTTCA TGTGCATCAT GGTGTTGGTA GATATTTAGG TGTGAGACG	2820
15	CTCGAAGTGG GGCAAACGCA TCGTGATTAT ATTAAATTGC AATATAAAGG TACGGATCAA	2880
	CTATTTGTTT CAGTAGATCA AATGGATCAA GTTCAAAAAT ATGTAGCTTC GGAAGATAAG	2940
	ACGCCAAAAT TAAATAAACT CGGTGGCAGT GAATGGAAAA AAACAAAAGC TAAAGTTCAA	3000
20	CAAAGTGTTG AAGATATTGC TGAAGAGTTG ATTGATTAT ATAAAGAAAG AGAAATGGCA	3060
	GAAGGTTATC AATATGGGGA AGACACAGCT GAGCAAACAA CATTTGAATT AGATTTTCCA	3120
25	TATGAACTTA CGCCTGACCA AGCTAAATCT ATCGATGAAA TTAAAGATGA CATGCAAAAA	3180
	TCGCGTCCAA TGGATCGCTT GCTATGTGGT GATGTTGGTT ATGGTAAAC TGAAGTTGCA	3240
	GTGAGAGCAG CATTCAAAGC TGTAATGGAA GGAAAGCAGG TTGCATTTTT AGTTCCTACA	3300
30	ACTATTTTAG CTCAGCAACA TTATGAGACG TTAATTGAGC GTATGCAAGA TTTTCTGT	3360
	GAAATTCAAT TAATGAGTCG TTTTAGAACG CCTAAAGAGA TAAAACAAAC TAAGGAAGGA	3420
	CTTAAACTG GATTTGTTGA CATAGTTGTT GGTACACACA AATTACTTAG TAAAGATATA	3480
35	CAGTATAAAG ATTTAGGGCT GTTGATTGTA GATGAAGAAC AACGATTGG TGTACGCCAT	3540
	AAAGAGCGTA TTAACATT AAAACATAAT GTAGATGTAC TAACATTGAC TGCAACCCCA	3600
	ATAGCTAGAA CATTGCATAT GAGTATGCTA GGTGTGCGG ATTTGTCAGT GATTGAAACG	3660
40	CCGCCAGAAA ATCGTTTCCC AGTTCAAACA TATGTATTAG AACAGACAT GAGTTTTATC	3720
	AAAGAAGCTT TAGAAAGAGA ACTATCCCGT GATGGCCAAG TGTTTTATCT TTATAATAAA	3780
45	GTGCAATCCA TTTATGAAA ACGAGAACAA CTCCAGATGT TAATGCCAGA TGCTAACATT	3840
	GCAGTTGCTC ATGGACAAAT GACAGAGCGC GATTTAGAAG AAACGATGTT AAGTTTTATC	3900
	AATAATgAAT ATGATATTTT AGTAACGACG ACGATTATTG AAACAGGTGT CGATGTCCCA	3960
50	AATGCAAATA CTTTGATCAT TGAAGATGCA GATCGCTTTG GATTGAGTCA GTTGATCAA	4020
	TTAAGAGGTC GTGTTGGTCG TTCAAGTCGT ATTGGTTATG CATACTTCTT ACATCCAGCA	4080
	AATAAGGTAC TAACTGAGAC TGCAGAAGAT CGATTACAAG CGATTAAAGA ATTTACGGAG	4140

55

TTAGGTAAAC AACAGCACGG CTTTATTGAT ACAGTTGGAT TTGATTGTGA CAGTCAAATG 4260
 TTAGAAGAAG CTGTAAATGA AAAACGTGGT ATTAAGGAAC CAGAATCTGA GGTGCCAGAA 4320
 5 GTCGAAGTTG ATTTAAACTT GGATGCATAT TTGCCAACAG AATATATTGC AAATGAACAA 4380
 GCTAAAATTG AA 4392

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 729 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

20 TTTCTTTTGA ATCTATATCG AGGTGGTTGG TAGGTTTCATC TAAAATAAGT ACATTGTCAC 60
 GTTGCAACAT AAGTAGTGCT AGTTGTAAAC GTGCTTTTTTC ACCACCAGAT AAATCATTTAA 120
 TTATCTTTTTT AACATCGTCT TGTACAAATA AGAAACGTCC AAGAACTGCT CGAATATCTT 180
 25 TTTCATTCAT TAACGGATAT TGATCCCACA CATAATCTAA AATCGTTTTTA CTAGATTTAA 240
 ATTCTGCTTG CTTTTGATCA TAATAACCAA TTTGTAAATT TGCGCCGAAA GTAATATCGC 300
 30 CATTAAAGCGC TTTTGTGTA TTAGCAATAG TTTTAATTAA GGTCGATTTT CCAATACCAT 360
 TTGGCCCAAT GATTGCTATA TGATCGCCTT TAGAGACCTC TATACTCATA GGTTTGGTAA 420
 TTGCAGTTTG ATAACCGATT TCTAAATTTT TTACATGCAT GACGTCATTA CCTGTATTCC 480
 35 GGTCAAAGCC AAATTGAATA TTTGCACTTT TGGCATCTAA CATTGGTTTA TCAATGCGTT 540
 CCATTTTTTC TAAAATCTTA CGTCTACTTT TTGCCATTCC ACTTGTTGAA GCACGGGTAA 600
 TATTTTTCTC AACAAAAGTT TCTAATCGTT TTATTTCTGC TTGTTGACTT TCATATTCTT 660
 40 GCATTCGTTT TTGATAATAT AAATCCCGTT GCTGTATAAA TTCCTCGTAA TTACCAACAT 720
 AGCGTTTGA 729

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13856 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

	5	10	15	20	25	30	35	40	45	50
1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	42	43	44
45	46	47	48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63	64	65	66
67	68	69	70	71	72	73	74	75	76	77
78	79	80	81	82	83	84	85	86	87	88
89	90	91	92	93	94	95	96	97	98	99
100	101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120	121
122	123	124	125	126	127	128	129	130	131	132
133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154
155	156	157	158	159	160	161	162	163	164	165
166	167	168	169	170	171	172	173	174	175	176
177	178	179	180	181	182	183	184	185	186	187
188	189	190	191	192	193	194	195	196	197	198
199	200	201	202	203	204	205	206	207	208	209
210	211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230	231
232	233	234	235	236	237	238	239	240	241	242
243	244	245	246	247	248	249	250	251	252	253
254	255	256	257	258	259	260	261	262	263	264
265	266	267	268	269	270	271	272	273	274	275
276	277	278	279	280	281	282	283	284	285	286
287	288	289	290	291	292	293	294	295	296	297
298	299	300	301	302	303	304	305	306	307	308
309	310	311	312	313	314	315	316	317	318	319
320	321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340	341
342	343	344	345	346	347	348	349	350	351	352
353	354	355	356	357	358	359	360	361	362	363
364	365	366	367	368	369	370	371	372	373	374
375	376	377	378	379	380	381	382	383	384	385
386	387	388	389	390	391	392	393	394	395	396
397	398	399	400	401	402	403	404	405	406	407
408	409	410	411	412	413	414	415	416	417	418
419	420	421	422	423	424	425	426	427	428	429
430	431	432	433	434	435	436	437	438	439	440
441	442	443	444	445	446	447	448	449	450	451
452	453	454	455	456	457	458	459	460	461	462
463	464	465	466	467	468	469	470	471	472	473
474	475	476	477	478	479	480	481	482		

	AAATGACTTT TAACATTATC TCTAGAGCCA TCTTTTAAAG ATGTTTGAAT AACGTCATAT	1920
	GTGATATCTG CAAATGCAGT TGTTTTATCT TTTACTAATTT GCTTATTTTC ATAAGGATCT	1980
5	GATATTTTAT CAATGTGCTT GTCATCTTTT TTAATATCAT CTAACGTTTT CTTAATATCT	2040
	TTAGTAATGT TCGGTTGCAC AATACCATCA TCTTTAGTCG TCTTAAAGAC AACACGTATT	2100
10	TGTGCCTTTT CACTATCTTG ATTAAAATGT TTTTCAATCT TTTTATTCGT ATCTAACGAC	2160
	TCTAATCCTG TCAITTTAAT ATCATTGTCA AATTTTCGGT CATTGTAGC AAGTGGTATC	2220
	AATATTGCAG CTACAATCAC TATCCATGCA ATGACCGCGG ACCATTTATG TTTTGCATG	2280
15	AATGTCCCCA TCTTATATAA AAATTTTGCC AAAGTATATT GCCTCCTTTT AAAATCAACG	2340
	TTATAGTTTA AATATACAGT GTAGATTATT GTTCGATTAT AGTATCTATC CCCGACCTCT	2400
	TAAAGAATCA ATTGGAAAAT TTTGTATATT AAACATACACA CAAAGGAGAA ATGTAGATGA	2460
20	AAGAGACTGA TTTACGAGTT ATAAAGACAA AAAAAGCATT GTCGAGTAGC TTGCTACAAT	2520
	TGTTAGAACA GCAATTATTC CAAACGATTA CTGTCAATCA AATTTGCGAC AACGCACTCG	2580
	TACACCGTAC AACATTTTAT AAACATTTTT ATGATAAATA TGATCTTCTA GAGTACTTGT	2640
25	TCAATCAATT GACTAAAGAC TACTTTGCTA GAGATATCAG TGACCGTCTT AATCATCCAT	2700
	TCCAAACGAT GAGTGATACG ATTAATAATA AAGAGGATTT GAGAGAAATC GCAGAATTCC	2760
30	AAGAAGAAGA CGCTGAATTT AATAAAGTAT TAAAAAATGT CTGCATTAAA ATTATGCATA	2820
	ACGATATCAA AAATAATAGA GACCGTATCG ATATTGACAG CGACATCCCA GATAATCTCA	2880
	TATTTTATAT TTATGACTCG TTGATTGAAG GTTTTATACA TTGGATAAAA GATGAAAAAA	2940
35	TTGATTGGCC TGGCGAAGAT ATTGATAACA TTTTCCATAG ATTAATCAAT ATTAAGATTA	3000
	AATAGTAGAT GAGAAACTCA TGAGCGTTAC CAACATTCAT AATAAAAACG ATAGTGKACA	3060
	CGTTAATGAA TTCGTGTACT ACTATCGTTT TTTATTTTTA TCGTGCTTAT CGCTATTAAA	3120
40	ACAACTGATA CACAACACAT AAACATGAA GAAAAAATA AATCCGCTAT CTAAATGACT	3180
	TTGACTCAGT TGTTTAAATG ACCAAATTGC TAATACAATT CCCATTATTA TTGAAATAAC	3240
45	GTATCTCACA TTCTTATACC TATAATCCTT TTCTAAAAT ATGGTTGCTA TTACTTAATT	3300
	TTTAAAGTTA TAAATAAAAA GAGCCAACCG CAATGGATGG CCCTTGTTCA TTATGAAGCA	3360
	TTAGAACATT TCTGAAACAA CCTTTTGTTT TAAGAAGTGT AATAAGTAGT CTGGACTACC	3420
50	TGTTTTAGCG TCCGTACCTG ACATTTTGAA ACCACCAAAT GGATGGTATC CAACAACTGC	3480
	TGAAGTACAG CCTCTGTTAA GGTATAAATT GCCTACATCA AATTCGTTTA CCGCTTTAAT	3540
55	CCAATGCTCG CGATTATTTG TAATCACTGC ACCAGTTAAA CCGTAATCTG TATCATTGTC	3600

	TTCTTCTTGC ATGATTCTAT CTTTAGATTT AAGTCCTGAA ATGATTGTTG GTTCTACAAA	3720
	GTAACCTTTT GAATCATCAG TGCCGCCACC TTGTTCTAAT TTACCTTCTT CTTTACCAAT	3780
5	CTCAATATAA TTTTAAATCT TATCAAATTG TTTTTTATTA ATAAGTGGGC CCATATACGT	3840
	ATTGTCTACA GTATTGCCCA ACGTTAATTC TTTTGTTAAT TTGATTGATT TCTCTAATAC	3900
10	TTCGTCATAA ACGTCTTTAT GCACAATTGC ACGTGAACAT GCTGAACATT TTTGACCAGA	3960
	AAAACCAAAT GCTGACGTTA CAATAGCTTC TGCTGCCATA TCTGTATCAA TATTTTCATC	4020
	AACTACAATG GCATCTTTAC CACCCATTTT AGCGATAACA CGTTTCAAGA AGTTTGTACC	4080
15	TTCTTGAACA ACGGCACTAC GTTCATAAAT TCTAGTACCT GTCGCACGTG ATCCTGTAAA	4140
	TGTAACGAAA TCGGTATCTT TATGATCAAC TAAGTAATCA CCAATTTCTT TCGGATCACC	4200
	AGGAACAAAG TTAAGTACGC CTTTTGGTAA TCCTGCTTCT TCTAAAATTT CCATTAATTT	4260
20	ATAAGCGATA TAAGGTGTAT CCTCAGCAGG TTTCAATAAC ACTGTATTAC CTGCCACAAC	4320
	TGGTGCTAAA GTTGTACCAG CCATAATCGC AAACGGGAAG TTCCACGGCG GAATGTAAAC	4380
	ACCTGTACCA ATTGATTAT AGAAATATTT ATTGTGTTCA CCTTCACGAT CAAGTACTGG	4440
25	CTTACCTTGA GCCAAGTCCA TCATTGAACG TGCATAGTAT TCAATAAAAT CAATACCTTC	4500
	AGCTGCATCA CCAACTGCTT CATCCCATGG CTTACCTGCT TCATAAACCA TAATTGCTGC	4560
30	AATTTCCGCT TTTCGACGAC GAATAATTGC CGAAACACGT AACATAAGCT CTGCACGATC	4620
	ATTTGCTGAC CATGTTTTCC AAGATTTATA AGCTTCGTTT GCTGCTTTAA ACGCATCTTC	4680
	AACATCTTGT TTTGTTGCCT TTGATGCATT TGCAATCACT TGTGATGTGT CTGCAGGATT	4740
35	GATTGATTTA ATTTTGTGAT CTTTGAAAAT CTTCTCTCCA TTAATCACTA ATGGTATGTC	4800
	TTGACCTAAT TCTTTTTTCCA CGTCTTTCAA TGCTTTCTTA AACATATCCA CATTTTCTTG	4860
	GACTGAAAAA TCGTAACCAG GTTCATTTTT AAATTCTACT ACCATGTACA CTTACCCCTT	4920
40	ATAAATTTTG AAAGTGGTTT AACCCTTTGA TTTAATGATA TAACATCATT TAAACTCATT	4980
	TTACTATGAT TAAGGTAGT TTTGCAATCG CTTTCATTTT TATGTTTTAT CACTTATTCT	5040
45	CAAGTATTTT GAAATTGATT GGTACTTTT TAAAATTAT ATGGGTCGCA ACTGCTACTT	5100
	TATCGTTTCG TCATTTAATG TTTCGGATGG TAGGTCATTA TCAATTTTAC GAACGACTTT	5160
	ACAAGGGTTT CCAACCGCTA AGCTGTGTGG CGGAATATCT TTAGTGACAA CACTACCAGC	5220
50	ACCAATCACA CTGCCTTCTC CAATCGTCAC CCCTGGTAAC ACGGCTACAT GACCGCCAAA	5280
	CCAAGTATTA CTGCCAATAT GAATGGGTCC GGCTTTTTCA AAACCTTCAT TTCTATGATG	5340
	GAAATTAAGT GGATGTGTCG CTGTGTAGAA TCCACAATTA GGTCTATATA AAACATTATC	5400

55

	TCCTAGTTTA	ACGTTCCAAC	CATAATCTGT	ATCAAAAGGA	ATCGAAATAC	TTACATTGTC	5520
5	TGTTGTTGTT	TGAAATAATT	GATCAATTAA	TTCTTTTCTT	TTATTTGTAG	CACTCGGTCT	5580
	TGTATGATTT	AATTCAAAGC	AAATATCTTT	CGCTCGTGCA	CGTTCATTGA	TTAAGTATTG	5640
	ATCAAAGTTT	GCATCGTACC	ATTTTTCTGC	TAACATTTTT	TCTTTTTCAG	TCATTACACC	5700
10	TTTCAACTCC	TAATAACTTA	TTACTTGTT	TAAAAGTTAA	TCAAATAAAC	CTTCGCCTAT	5760
	GCAACTAATA	CGCTATAACA	TTATGAAATC	ATGACCTTAT	CACCCTTATC	TATACAATTC	5820
	TCGCATCAAA	TACTGCTAAA	GTAGTAGATA	AATTCAATAC	TACAGACGCA	TTCATTTTTT	5880
15	AATCTATTAA	CGTACAATGT	GAGTAAGAGA	AATATAAAGG	AGTATGATAG	CGATGAGAAT	5940
	ATTAATTACA	GGCACAGTTG	CTATCTTAAT	CATTCTAGGT	TTGGTCAAAA	CGATACAAGA	6000
	TTACGAAATG	ACAAACGACA	CGAGTCGTcA	GTTGTCAGAC	AACAAAGATG	ATGATAAAGT	6060
20	CATCCATCTT	AATAATTTTA	AAAATTTACA	TGCGAAAGAA	TTTAACCCAT	CTGATTTCTT	6120
	TTAAGTCACC	TAAGAATTGC	AAATCCAGAA	GTCATTTAAG	TTTTACCTTT	CATTCATACA	6180
25	TCCTTTAATA	TTAATTACGA	CTTCTTTTAT	ATAGATGCTA	AGTAGAGAGA	TTGTTGTGCA	6240
	ATGTTTGAC	GGCAATCTCT	CTTTTTCTTT	TTAAAATTGG	TAAAAGTAAA	ACGCAACGAT	6300
	TGACTTATAT	ACCTATAGGG	GGTACATTAG	ACGTGTAACA	ATGAATCACA	GGGAGGCAAT	6360
30	AATGTGGCTA	ATACGAAAAA	AACAACATTA	GATATCACTG	GTATGACTTG	TGCCGCATGT	6420
	TCAAATCGTA	TCGAAAAGAA	ACTGAATAAA	CTTGATGACG	TTAATGCCCA	AGTGAATTTA	6480
	ACTACAGAGA	AAGCAACTGT	TGAGTATAAC	CCTGATCAAC	ATGATGTCCA	AGAATTTATT	6540
35	AATACGATTC	AACATTTAGG	TTACGGTGTC	GCTGTAGAAA	CTGTCGAATT	AGACATTACA	6600
	GGTATGACTT	GTGCTGCATG	CTCAAGCCGT	ATTGAAAAAG	TGTTAAATAA	AATGGACGGC	6660
	GTTCAAATG	CAACGGTCAA	TTTAACAACA	GAGCAAGCTA	AAGTTGACTA	TTATCCTGAA	6720
40	GAAACAGATG	CTGATAAACT	TGTCACTCGC	ATTCAAAAAT	TAGGTTATGA	CGCGTCTATT	6780
	AAAGATAACA	ATAAAGATCA	AACGTCACGC	AAAGCTGAAG	CGCTACAACA	TAAATTGATT	6840
45	AAGCTTATCA	TATCAGCAGT	ATTATCTTTA	CCACTATTAA	TGTTAATGTT	TGTACATCTT	6900
	TTCAATATGC	ATATACCAGC	ACTATTTACG	AATCCATGGT	TCCAATTTAT	TTTAGCTACA	6960
	CCTGTACAAT	TTATTATTGG	ATGGCAATTT	TATGTAGGTG	CTTATAAAAA	CTTAAGAAAT	7020
50	GGTGGCGCCA	ATATGGATGT	ACTTGTGTGCT	GTTGGTACAA	GTGCAGCATA	TTTTTACAGT	7080
	ATTTATGAAA	TGGTTCGTTG	GCTAAATGGC	TCAACAACGC	AACCGCATTT	ATACTTTGAA	7140
55	ACAAGCGCCG	TACTAATTAC	CTTAATCTTA	TTCGGTAAGT	ATTTAGAAGC	TAGAGCGAAG	7200

TTAAAAGATG GTAATGAAGT GATGATTCCCT CTAAATGAAG TACATGTTGG AGATACACTT 7320
 ATCGTTAAAC CAGGTGAAAA GATACCTGTT GATGGCAAAA TTATTAAAGG TATGACTGCC 7380
 5 ATCGACGAAT CTATGTTAAC AGGTGAATCT ATCCCTGTTG AGAAGAATGT TGATGATACT 7440
 GTAATTGGTT CAACGATGAA CAAAAACGGT ACTATTACTA TGACAGCAAC AAAAGTTGGC 7500
 10 GGGGACACTG CGTTGGCAAA TATTATTAAA GTTGTGCAAG AAGCTCAAAG TTCTAAAGCG 7560
 CCGATTCAAC GATTGGCAGA TATTATTTCT GGTATTTTCG TTCCTATCGT TGTGTTATC 7620
 GCACTATTAA CATTTATCGT GTGGATTACT TTAGTTACAC CAGGTACATT TGAACCTGCA 7680
 15 CTTGTTGCGA GTATTTCCGT TCTCGTCATT GCTTGTCCAT GCGCATTGGG ACTTGCTACA 7740
 CCAACTTCTA TTATGGTAGG TACTGGTCGC GCTGCTGAAA ATGGTATTTT ATTTAAAGGT 7800
 GGCGAGTTTG TTGAACGCAC ACATCAAAT GATACCATCG TTTTAGATAA GACGGGTACC 7860
 20 ATTACAAATG GTCGTCCAGT CGTGACAGAT TATCATGGTG ACAATCAAAC GCTACAATA 7920
 CTTGCTACTG CTGAAAAAGA TTCTGAACAC CCATTGGCAG AAGCCATTGT CAATTATGCA 7980
 AAAGAAAAGC AATTAATATT AACTGAGACA ACAACATTTA AAGCAGTACC TGGCCATGGT 8040
 25 ATTGAAGCAA CGATTGATCA TCACCATATA TTGGTTGGTA ACCGTAAATT AATGGCTGAC 8100
 AATGATATTA GCTTGCTTAA GCATATTTCT GATGATTTAA CACATTATGA ACGAGATGGT 8160
 30 AAAACTGCTA TGCTCATTGC TGTAAATTAT TCATTAAGTG GTATCATCGC AGTGGCAGAT 8220
 ACTGTCAAAG ATCATGCCAA AGATGCTATA AAACAATTGC ATGATATGGG CATTGAAGTT 8280
 GCCATGTTAA CTGGCGATAA TAAAAACACT GCTCAAGCCA TTGCAAAACA AGTAGGCATA 8340
 35 GATACTGTTA TTGCAGATAT TTTACCAGAA GAAAAAGCTG CACAAATTGC GAAACTACAG 8400
 CAACAAGGTA AGAAGGTTGC GATGGTTGGT GACGGGTGTA ATGATGCACC TGCATTAGTT 8460
 AAAGCTGATA TCGGTATCGC CATTGGTACA GGTACAGAAG TTGCCATTGA AGCAGCTGAT 8520
 40 ATTACTATTC TTGGTGGCGA CTTGATGCTT ATTCCTAAAG CCATTTATGC AAGTAAAGCA 8580
 ACCATTGCTA ATATTCGTCA AAATCTATTT TGGGCATTCT GCTATAATAT TGCCGGTATC 8640
 CCTATAGCTG CATTGGGCTT ACTTGCGCCA TGGGTTGCTG GTGCTGCAAT GGGACTAAGT 8700
 45 TCAGTAAGTG TTGTCACAAA CGCACTTAGA TTGAAAAAGA TCGGATTAGA ACCACGCCGT 8760
 AAAGATGCCT AGATTCTTAA ATAATGAAGG ATTCGTTGGT GATTCTGAGA TAGGCTAGTG 8820
 50 ATTGGCTCTA TAATGTCGCG GTTTAyaGt GGATCTTCGC TCCAAGTGCA TATATAGTna 8880
 CACTTTTCGC TTGGCGAATT AGTGATCTT ACCTAATAGc TCCGCCTATT AGGTTCCATC 8940
 ATTATTATAA ATAATAAGTA CACTACGGtT TACAGTTGGA TCTTCGCTCC AACTGCATAA 9000

55

GAAATTTTAA ATGTTGAAGG TATGAGCTGT GGTCACTGCA AAAGTGCTGT TGAATCTGCA 9120
 TTAATAATA TTGACGGTGT CACTTCAGCT GACGTTAACC TTGAAAATGG TCAAGTAAGT 9180
 5 GTTCAATATG ATGACAGTAA AGTTGCTGTA TCTCAAATGA AAGACGCAAT TGAAGATCAA 9240
 GGTTACGATG TCGTTTAATT AGGCAATATT CAACGTCATC AACACCAAAT TAAAAAATCG 9300
 10 AACTGATGAG AATCCCAACA ATCCAAATTA TCTCATCAGT TCGATTTTTA ATTTACTCGT 9360
 AACCTAGTAT CTCCAGTCTG CAATACATCT AATGTTGCAT CTAATGCATC GACAATTAGA 9420
 TTTTAACTG CAGCTTCAGT ATAAAACGCA ATATGTGGTG TTAATATGAC ATCTTCCCTG 9480
 15 TCAATCAACG ATTCTAACAA TGGATCGTTC AGTGTTTTGC CCCTTTGATC ACTTGGGAAA 9540
 AGTTTGC GTT CAAATTCATA CGTATCAAGT GCTGCACCTT TAATCACACC ATTGTCTAAT 9600
 GCGTCTAATA ACGCCTTAGT ATCTACTAAA GAACCTCTCG CACAATTGAC AAATACTGCG 9660
 20 CCCTTTTTAA AATGTTTAAA TAATTCAGCA TTAAATAGAT AATGATTATA TTTCGTTGCA 9720
 GGTACATGTA ATGTCACGAT ATCAGCACCT TCAACCGCTT CCTCAATCGT ATCTTTGTAA 9780
 TCGACATACG TTGCAATTTT AGCATTAGGA AACGGTCGTA TGCGACCACA TCACTTTGAT 9840
 25 AACCATTGGC AAATATATCG GCTACTACAC GGCCAATTTC ACCTGTACCA ATAACAGCTA 9900
 CTTTTAAATC TTTAATGGAT TTCGATAAAA TAGTAGGTTT CCATCTAAAA TCATGCTCCC 9960
 30 GCACTTTCGT TTGAATTTGA TTAAAATGAC GAACCACATT AATAGCCTGG TTCACAGCAA 10020
 ACTCCGCAAT TGAATTCGGA GAGTATGACG GCACATTTGA CACAATAAAG TTATACTTGT 10080
 TTGCTAACTC CAAATCATAT GTATCAAATC CAGCACTACG TTGTGCGATT TGTTTAATAC 10140
 35 CTAGTTCATT TAATCGTTTA TAAACATGCT CTGATAATGG TATTTGTTGT GATAGCGATA 10200
 AGCCATCATA ACCAGCGACA CCTTCAACAT TGTCATCAGT TAATGCTTCT TTAGTAATAT 10260
 CTACCTCAAC ATGATGTTTC TCTGCCCACG CCTTGATATA AGGCATATCT TCATCACGTA 10320
 40 CACTCATGAT TTTAATTTTT GTCATTTTAA CATCACCTT AACTTTATTA TTCATATAAA 10380
 TATGCTAGTT CTGTTAATCT TATTGCAGCT TCGTCTAATT TCTGGTCATC TAACGCCAAT 10440
 GAAATTCTCA CATAACGATT ACCATTCTCT CCAAATGGTT TCCCTGGAGC AACAAGTATT 10500
 45 GACTTCTCTT GCACTAAAAA TTGCTCAAAT TGCTCGCTGT CATAACCAGG CCGTGTTTCC 10560
 AACCATACAT ATATGCCACC TTTAGCATGA ACAAATGGCA AATCAGCTTT TGCAAGCATG 10620
 50 GCTTCGAATC GGTCAACGACG TGTTTTAAAT ACATTGCTTT GTTCTTCTAA AAAATCATCA 10680
 TAATGATTCA AAGCATATAT TCGGCATCT TGTAATGCAC CAAACATCCC AGCATTTGTG 10740
 TCGGTTTGGT ACTTTTTCAA AGCTTGAATC ATATCTTTAT TACCAACTGC AAAACCGACT 10800

55

	CCATTTTCCG	AAGCAAGTAT	ACTAGGATTT	TTAGCGTCGA	AACCGAAAGC	ACCATAAGCA	10920
	AAATCATGCA	CGATTTTAGT	GTCTGTACCT	TTAAATTTAG	cTATCGCTTC	ATCAAAAACT	10980
5	TCTTTCGTAG	CTGTCGATCC	AGTTGGATTA	TTTGATACG	TTAAATAAAT	GAGTTTGTGTT	11040
	TTATCTATTA	TTTGTGAATC	AACTTTGGAC	CAATCTGGCA	AATAATGTGG	CGGTTCTAAA	11100
10	TTAAGCGGGA	CTGGCTTGCC	ATCAGCTAAA	AGTACACCTG	CTAAATAATC	CGTGTAGCCT	11160
	GGATCAGGTA	GTAATACATA	GTCTCCTGGA	TTGATAACAC	ATGTTGGTAC	TGCCACTAAT	11220
	CCATTTTTTG	TACCATATAA	AATGCATACT	TCATCTTCTT	TATCTAACGT	CACATTATAT	11280
15	TGTCTTTGAT	AAAAATCTAC	AATAGCTTGC	TTGAACGCTT	CTTTACCATG	AAAAGCACCA	11340
	TATTTTTGAT	TTTCAGGAAT	AGTTAGTGCT	TTTGTAAAAT	GATCAATAAT	ACCTTGTGGC	11400
	GTGGGCCCCAT	CAGGGATTCC	AACTGCCATA	TTAATTAATG	GCAATGGTCC	ATGTTGCGATT	11460
20	TTACGTCCCA	TCGTTTTCCC	GAAATAACTA	TCAGGGATAT	TTGCTAATTT	GTTAGAGATC	11520
	ATCAAATTCC	TCCTCTATCA	TTAAACATAG	CCTGGGCGAC	TATCATAATC	CTAACAACCTT	11580
	GTATCACTCT	CATTTAGATG	GTTACAATGA	CATCGCCATT	CACCGTTATG	TTCAACAGAA	11640
25	CTTATGACAC	ACGTTGTATT	GAATGAATTT	ATTTTCATTT	TAGGTAGGTA	TAATATTATT	11700
	GTCAATATTA	GGAATTTTCA	GATTAATATG	CACTCAATCG	TTATGATTTA	ACTGTCATGC	11760
30	ATATCCGCAT	GCGCAACCAG	TTAGATATGC	TTATATAAAG	TATAACGCCC	ATCAAGGTAC	11820
	GTATTCAAAC	GTGAACCTTA	ACAGGCGTCA	TTCATTGTTA	AATAAAACTT	CTTAAGCACA	11880
	TACTTATTTT	ACTATGCCTT	TTACGTTCCC	CTTATACTTT	TCTCACATCT	TTCTCTTAGA	11940
35	CTACTCCCTT	ATACGCCCCG	CTCAATATCT	TTAATCATTT	CATCTACAGT	TATTTTCGCA	12000
	CTCGTTAAGA	CAATAGGAAC	GCCTGCACCT	GGATGCGTAC	TTGCACCTGC	AAAATATAAA	12060
	TCTTTATAAT	CTCGCGATAC	ATTTTGTGGA	CGATAATAAT	TACTTTGCGC	TAAAGTTGGC	12120
40	ATTAAACCGA	ATGCCGAACC	AAATTTGCGA	TGATACGTTT	GCTCAAAATC	ATTTGGCGTA	12180
	AAGATTGTTT	CTGAAACAAT	ATGCGATTTT	ATATCTTCAA	ATACTTCAAT	CGTTGCTAAT	12240
45	TTACGATAAA	TAATTTCCCT	TATTTGTTGC	GTCAAAGCTT	CATCTGACCA	ATCGATTCCG	12300
	CTACCTGTTT	TAAGTTCCGG	CGTCGGCATT	AGCACATAAA	TACCAGTTTT	GCCTTCTGGC	12360
	GCAAGTGATT	TATCAGCGAC	CGCTGGTACA	TACACATAAA	TAGAAGGATC	ATATGATAAA	12420
50	CGTCCCTCAA	ATATTTCTTC	AATATTGCCT	CTAAAGTCAT	CTGAAAAAAT	AACATTATGA	12480
	AGTCTCACTT	GATCTGTCAC	ATCAATATCT	ATACCGATAT	ACATTAAAAA	TGCTGAACAA	12540
	GAGTAATCTA	AGTCTGCAAT	TTTATGTGGT	GGATACTTTT	TAATAGGTGC	AAAATCTGGC	12600

55

5 ATGTCACCAT TCACTTTTAT CGCATCGGCC CGTTTGAATT TAGGATCAAT AATAATTTGC 12720
 TCAATTTTCAG CATTTAGTTC AATATTAACG CCTAAGTCTT TATTTAATTG CGCTAGcCCT 12780
 TGAGCCATGC CATACATACC GCCTTTAATA AAATGCACAC CAAACATCAT TTCAATCATA 12840
 GGAATAATTG AATATAGTGA CGGGCCTCGT TTTGGATCAA TTCCTATGTA TAACGTTTGA 12900
 10 AACGCTAAAA GCTTTTGTAT CTTTTCGTTA TCAATATAAT GTTCAATTAG CTGATCTGCA 12960
 TGATTTAACG TTTTAACTT AGCACCTTGC ACAAGTGACG TCATATTATA AAAGTCACTC 13020
 GGTTCGCGAT ACGTTCCTTC TAAGAAATAG CGACGTGCAA TTTCATATTT TTTATAAACA 13080
 15 TCCGTTAAAA AGGACATAAA ACCATGCGTT GAACCAGGTT CTATACTTTC TAGCATTTCG 13140
 TGTAATTCAG CTAATCTGT AGGCACCGTT ATACGATCAT CGTGGTCAAA ATACACATCG 13200
 TAAATATAAC GTAATTGTCT CAATTCAATA TAATCTTCAT AATTTTACC ACACGCTGTA 13260
 20 AAAACATCTT TATAAACATC TGGCATCATG ACAATTGTGG GACCCATATC AAATGTAAAG 13320
 CCGTCTTTCT TTAATTGATT CATACGCCCCG CCTACATTAT TATTTTTTTC AAATATCGTC 13380
 ACTTCATGAC CTTGAGAAGC AATACGGGCT GCCGCTGCTA ATCCTGTGAC ACCTGCACCA 13440
 25 ATTACTGCAA TCTTCATTAT TCAACCACCT ATATTCTATG ATATTTACTA TTTATTTTAT 13500
 GAAACAACCT TGCCTTTTTT CTCTTATCCA CAAAAACACG TTCATGTAAT GTATAGTTAG 13560
 CCTGTCTCAC TTCGTCCAGT ATTTCAATAT ATATACGTGC TGCTAATTCT ATGATTGGTT 13620
 30 GTGCTTCAAT ACTAAATACT TTGATTTGAT CCATAACATC TTGAAAATCT TTTTCTGCCA 13680
 TAGCTGCATA ATATTCCTCAT AAGTCAATAT AATGATTATT AACACCATTT TGGTACACTT 13740
 35 CAGCAATATC AACTTCATAT TGCTTTAATC GTTGCTTACT AAAATATATC CGTTCATTGT 13800
 CAAATCTTC ACCGACATCT CTTAATATAT TAAAGGATC CTCTAGAGTC GACCTG 13856

(2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10088 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

50 ATATATAAAT ATAGATTAAG TATATAGATT AATCAACTTT TTTGGAAGAG CAAATCACGC 60
 AATCAACAAA TAATATAAGA AGTTTTTGCG ATAGTTTAA AATAGCTGTA ATAGAATACT 120
 AAATGTGACA AACTTAGAAC TAATATCAAG TGTGATGTT TTGAATATAA AAATGCTAAT 180

	ATAATTGGTT AATATATGAG TAATTAGAAA ATAGACAAAG GATGACGATT TATGTATATC	300
	AATATGAAAG ATTATGGGTT AACAGGCATA AACAAAACTA AAGATACTCG AGCAATACAA	360
5	CGTGC GTTAA ATCGTGGAAG ATGTAAACCA ACGACAGTTT ATATACCGAA AGGGACGTAT	420
	GATATTTGCA AACCATTAAAC GATATATGGC AATACAACAC TTTTGTTAGA TAATGAAACT	480
10	ATTTTACGCC GATGTCATTC TGGTCCTTTA TTAAAAAATG GTCGTCGCTT TGGTTTTTaT	540
	CGTGGTTATA ATGGACACAG TCATATTCAT ATTAAAGGCG GCAAGTTTGA TATGAATGGT	600
	GTATCGTATC CTTATAACAA TACAGCTATG TGCATTGGGC ATGCTGAAGA TATTCAATTA	660
15	ATAGGTGTGA CCATTAAGAA TG TAGTGAGT GGTCATGCAA TTGATGCTTG TGGGATTAAAC	720
	GGACTCTATA TTAAAGCTG TTCATTTGAA GGATTCATAG ACTATAGTGG CGAACcTTT	780
	ATTCTGAAGC AATACAATTA GACATTCAAG TACCTGGTGC TTTTCCAAAA TTCGGAACgA	840
20	CAGATGGTAC GATAACGAAA AATGTCATTA TCGAAGATTG TTATTTTGGA CCTTCAGAAT	900
	TGCCCCGAAAT GGGAAAGTTGG AATCGTGCTA TTGGCTCACA TGCAAGTAGA CATAATCGAT	960
25	ACTATGAGAA TATTCATATT AGAAATAATA TATTTGAAGA TATACAAGGT TATGCATTAA	1020
	CTCCCTTGaA GTATAAAGAT GCTTTCATTA TTAATAATAA GTTTATTAAC TGTGaGGGTG	1080
	GCATTAGATA TTTAGGAGTT AGAGATGGTA AAAATGCAGC AGATGTGaTG ACAGGaAAAAG	1140
30	ACTTAGGTTTC CCAAGCAGGC ATAAATATGA ATATAATTGG AAATGAATTT AAAGGATCAA	1200
	TGTCTAAAGA TCGGATACAT GTACGTAATT ATAATAATGT TAAACATAAA GATGTATTAA	1260
	TCGTTGGGAA TACATTCAAT AATTCGACTC AATCAATTCA TTTAGAAGAT ATTGATACAG	1320
35	TGTTTTTAAG TCCTGTTGAA GCGGGTATTC AAGTTACTAC AATCAATGTA GATGAAATAA	1380
	AAAAGTAAAA AGTTTCGCAT GACATTAGGA TTAAGAATAG TAGATAATTT TTGAAAGCGC	1440
	ATTēATAAAA CGGTATAAAT ATGCTATAAT AAACCCAATT ATCTGATAAA AGGGGTATTT	1500
40	TGACGGTAAT GATAATACAA GATAGACAAC TTTCTATACT CTAATATAGT GAGTTGAAGT	1560
	AGCTTGT CAT AATCATCATG AGGGGGAAAT TTATGGCTTA TTTCAATCAA CATCAATCAA	1620
45	TGATATCGAA AAGGTATTTA ACATTCTTTT CAAAATCAAA GAAAAAGAAA CCGTTTAGTG	1680
	CGGGACAACT TATTGGACTA ATATTAGGTC CATTACTTTT CCTATTAACA TTATTATTCT	1740
	TTCATCCACA AGACTTACCT TGGAAAGGCG TCTATGTTTT AGCGATTACT TTATGGATTG	1800
50	CGACTTGGTG GATTACTGAA GCAATTCCTA TTGCAGCAAC GAGCTTATTA CCAATTGTGT	1860
	TATTACCATT AGGTCATATA CTTACACCAG AACAAAGTATC ATCCGAATAT GGCAATGATA	1920
	TTATCTTTTT GTTTTTAGGT GGATTTATTT TGGCAATTGC AATGGAAAGA TGGAATTTAC	1980

55

	TTGGATTTCAT GGTGGCAACA GGATTCTTAT CTATGTTTGT ATCGAACACT GCAGCTGTAA	2100
	TGATTATGAT TCCGATTGGT TTAGCAATTA TTAAGGAAGC ACATGATTTA CAAGAAGCCA	2160
5	ATACGAATCA AACAAGTATT CAAAAGTTTG AAAAATCTCT AGTTTTAGCA ATTGGCTATG	2220
	CAGGTACGAT TGGTGGCTTG GGTACATTAA TCGGAACCCC GCCATTAATT ATTTTAAAAG	2280
	GACAATACAT GCAACATTTT GGACATGAAA TTAGTTTTGC TAAATGGATG ATTGTAGGGA	2340
10	TTCCAACGGT CATTGTTTTG TTAGGTATTA CTTGGCTCTA TTTAAGATAT GTTGC GTTTA	2400
	GACATGATTT GAAATATTTa CCTGGTGGTC AGACGTTAAT TAAACAAAAG TTAGACGAGC	2460
15	TTGGCAAAAT GAAGTATGAA GAAAAGGTAG TACAACTAT CTTTGTACTT GCTAGCTTAT	2520
	TATGGATTAC AAGAGAGTTT CTTCTGAAAA AATGGGAAGT TACGTCATCT GTTGCAGATG	2580
	GTACGATTGC TATTTTTATA TCAATATTAT TATTTATTAT TCCAGCTAAA AATACTGAAA	2640
20	AACATCGCCG TATCATTGAC TGGGAAGTTG CAAAAGAGCT CCCTTGGGGT GTATTAATTT	2700
	TATTTGGTGG CGGTTTAGCA TTAGCGAAAG GTATTTCTGA AAGTGGTTTA GCAAAATGGT	2760
	TAGGCGAACA GTTGAAATCA TTAAATGGTG TTAGTCCGAT TCTTATTGTA ATTGTCATAA	2820
25	CAATCTTTGT CTTATTTTTTA ACTGAAGTGA CATCTAATAC TGCAACTGCA ACGATGATTT	2880
	TACCGATTTT AGCAACGTTG TCTGTTGCTG TTGGAGTGCA TCCATTACTA CTTATGGCAC	2940
30	CTGCAGCTAT GCGCGCTAAC TGTGCATACA TGTTACCAGT AGGGACACCA CCGAATGCAA	3000
	TTATCTTTGG TTCTGGTAAA ATATCTATCA AACAAATGGC ATCAGTAGGA TTCTGGGTAA	3060
	ACTTAATCAG TGCAATAATT ATTATTTTAG TCGTGTATTA TGTAATGCCT ATAGTTTTAG	3120
35	GTATTGATAT AAATCAACCA CTGCCATTGA AATAGTAATT GCAGATTAGA ACGAAAAATA	3180
	AAAGGTTACA TTAGCAATTG CTTGGACGAG TGGTAACGAA ACGTATACCG CAGCATCGTG	3240
	TAA SA ACAAT ACAAAACAAA GAAAGTCAAC CAAGGATGGA TTCCTATTTT AATCCTTGGT	3300
40	TGACTCTTTA TTTTATTTAA ATTGTAGAAC CTAGAAAATA AAGTTTAATT AAAAGCACCA	3360
	ATCATTTCTA CTTTGAAATC TAAGGTTTCT AAAATAGCAA TGACTTTCTT TATATCGGTT	3420
	GTAATTGCAG AATCAGCCTG AACGAAAAAT CGATACATAC CTAATTGTGT TTTTAAAGGA	3480
45	CGAGACTCAA TCCAGGATAA ATTAATATTA AACAAAGCAA ATGTATTAAG CACACTTGCT	3540
	AACAACCCAG GTTTATCATG CATTGGTGTA ATTAAAAACA TCAATGATGT CGCATTTTGA	3600
50	TCAAATTGCT GCTGATTTTT TATAACTAAA AAACGTGTCA CGTTATGTGG ATAGTCTTCA	3660
	ATATGTGTAT CAATAGGTGT AAAACCATAA GctTCGCCAC TACCTAAAGG TGCAATTGCT	3720
	GCAACGCCAT TTTCAATTTT AGTCAAACCT TGAATTGTAC TGTCGACATA ATCATAGTCA	3780

55

	TTTTTAATAT CAGAAATGGA ATCTGTTCCA TTACCATATA ATGCAAAGTT AATATCTAAA	3900
	CGTATTTTCAC CGTGTGCAAA GACATCTTGC TGTGCAAGTG CATCTGCCAC AATGTTGATT	3960
5	GTTCCCTTCTA TAGAATTTTC AATAGGGACA ACACCAATCG ATGTGTCATC ATCTGCAACT	4020
	GCCTTGATGA CTTCAAATAA ATTTGACTTT GGTGAAAAG TTGCTTCATT TTCAGAAAAA	4080
10	TACTGACGAC AAGCCAAATA TGAAAATGTA CCTTTAGGGC CTAAATAATA TAATTGCATA	4140
	TGCTACACCT CTACTAACTT AATGATGGAA AGGGCACTGG TTAGCATTTG ATTCTTTCTT	4200
	TTTATAGAAA AAGTTTGGAT CTTTTACTGT ATTGTCATAT CCGTGATGAT AATTTGACGT	4260
15	CAATGTTGGA GATAATGGCG GTGCTAGCCA AGACCATTTT CCGGTAACCT GACGACCTTG	4320
	TTGTGCTTCG TTACGTTTCA ATAGTTTCGA TTGCTTTGCA GCGGTCAAAT GATCGACAAT	4380
	TGATACGCCT TCTTTTTTAA AGGAATGATA CACAGCATAG TTCAATTCAA CAAGTGCTCG	4440
20	ATCTTTATTA AATGAATTAT TTTTAAGTGT ATCAAATTCA AACGCATCTG CAACTTTTTTC	4500
	TAGTAAATTG TAACGGTAAT CATCAATAAA GTTACGTACG CCAATTTTCA TTAGCATATA	4560
25	CCAACCGTTA AAGGGTGCG TTGGATATAC AATGCCACCG ATTTTIAAGT CCATATTGGA	4620
	AATGATAGGG ACTGCATACC ATTTTAAGTT CAATTTTCTT AATTTTGGAT AATGATTATG	4680
	TTCAATAGGT ACTTCTTTAA TTAATGAAGT AGGATATTCTG TAAAATTTAA CTGACTCATT	4740
30	AGGTAATTGG TAAATCAGTG GTAACACGTC AAAATTAGTA CCTTTTCCTT TCCAACCTAA	4800
	GTGATTGCT AAGCGTGTA CTTCTTTTTT AGCAGGATCA CCACAATTGT CATAGCCAGC	4860
	ATAGCGAATT AATTGATTGT TGAAAATTTT AGGTCCATCC TTTGGAGCAT ATATAGTAAT	4920
35	ATACGGCTTT AATTACCTT CATTTGTAGC CTGTGTAATA TGATAAGTAA TTGATGATAA	4980
	GAACGATGCT TCGTCAGTAA CATCTCTTGC ATCAATGACA TTTAACGAAT CCCAAAATAA	5040
	ACGACCAATG CAACGATTG AATTACGCCA AGCCATTTTA GCACCATAAA TAAGTTCTTC	5100
40	TTCTGTATGT GTATATGTCC CAGTTTCTTT TATTTCTAGT TCAATGTCAT GTAAACGTTT	5160
	ATTGATAATT TGCCTTTCAT AATGACACTC TTTATACATG TTTTCTATGA AAGCTTGAGC	5220
45	CTCTTTAAAT AACATTAACA ACACCTCGCT TTATATTATA GTCTACATTA TTAAAATACT	5280
	CTTAAAAATT ATGTATATGT CATTAAATTG TTGGTTGATT TTAATTAAAA GTATGGAAAT	5340
	TAAGGGGCTC TTATGTATAT AAAAAATGA ATTATGATAA AATGTAAGAA AATATTTAGG	5400
50	TCGATTGGAG AGATACAAGT GTACCAATTA GAAGACGACA GTTTAATGTT ACATAATGAC	5460
	TTATATCAAA TAAATATGGC TGAAAGTTAT TGGAATGATA ATATTCATGA AAAAAATGGCT	5520
55	GTATTTGATT TGTATTTTAG AAAAAATGCCA TTTAATAGTG GCTATGCTGT TTTAATGGT	5580

TTAAAGTCTA TTGGCTACAA GGATGATTTT TTATCATATT TAAAAGATTT AAAATTCACA 5700
 GGCAGCATCC GTTCGATGCA AGAAGGCGAA TTATGCTTTG GTAACGAACC ATTGTTACGC 5760
 5 GTAGAAGCAC CATTGATTCA AGCGCAATTA ATAGAAACAA TTTTATTAAA CATTGTAAAT 5820
 TTCCATACAT TAATTACAAC AAAGGCTAGC AGAATTCGTC AAATTGCATC AAATGATAAA 5880
 10 TTAATGGAGT TTGGTACACG TCGTGCGCAA GAAATTGATG CAGCATTGTG GGGCGCTAGA 5940
 GCTGCTTACA TCGGGGGCTT TGATTCTACA AGTAATGTTA GGGCGGGGAA ATTATTTGGT 6000
 ATACCTGTGT CTGGTACACA TGCACATGCA TTTGTCCAAA CTTATGGAGA CGAATATGTT 6060
 15 GCCTTCAAAA AATATGCTGA AAGACATAAA AATTGTGTGT TCCTAGTAGA TACATTCCAT 6120
 ACTTTAAAT CTGGCGTGCC AAATGCAATA AAGTTGCAA AAGAATTAGG TGACAAAATT 6180
 AACTTTGTAG GTATTCGATT AGATTCTGGA GATATCGCTT ATTTATCTAA AGAGGCAAGA 6240
 20 CGTATGCTTG ATGAAGCAGG ATTTACTGAA ACTAAAATTA TCGCGTCTAA TGATTTGGAT 6300
 GAAGAAACGA TTACGAGTTT GAAAGCACAA GGTGCAAAAG TAGATTCTTG GGGCGTTGGT 6360
 ACAAAGCTGA TTACAGGATA CGATCAACCA GCATTAGGTG CAGTATATAA ACTTGTAGCT 6420
 25 ATTGAAAATG AAGATGGTTC ATATAGTGAT CGTATTAAAT TATCAAATAA CGCTGAAAAG 6480
 GTTACGACGC CAGGTAAGAA AAATGTATAT CGCATTATAA ACAAGAAAAC AGGTAAGGCA 6540
 30 GAAGGCGATT ATATTACTTT GGAAAATGAA AATCCATACG ATGAACAACC TTTAAAATTA 6600
 TTCCATCCAG TGCATACTTA TAAAATGAAA TTTATAAAAT CTTTCGAAGC CATTGATTTG 6660
 CATCATAATA TTTATGAAAA TGGTAAATTA GTATATCAAA TGCCAACAGA AGATGAATCA 6720
 35 CGTGAATATT TAGCACTAGG ATTACAATCT ATTTGGGATG AAAATAAGCG TTTCTGAAT 6780
 CCACAAGAAT ATCCAGTCGA TTTAAGCAAG GCATGTTGGG ATAATAAACA TAAACGTATT 6840
 TTTGAAGTTG CGGAACACGT TAAGGAGATG GAAGAAGATA ATGAGTAAAT TACAAGACGT 6900
 40 TATTGTACAA GAAATGAAAG TGAAAAAGCG TATCGATAGT GCTGAAGAAA TTATGGAATT 6960
 AAAGCAATTT ATAAAAAATT ATGTACAATC ACATTCATTT ATAAAATCTT TAGTGTTAGG 7020
 TATTTAGGA GGACAGGATT CTACATTAGT TGGAAACTA GTACAAATGT CTGTTAACGA 7080
 45 ATTACGTGAA GAAGGCATTG ATTGTACGTT TATTGCAGTT AAATTACCTT ATGGAGTTCA 7140
 AAAAGATGCT GATGAAGTTG AGCAAGCTTT GCGATTCAAT GAACCAGATG AAATAGTAAC 7200
 50 AGTCAATATT AAGCCTGCAG TTGATCAAAG TGTGCAATCA TTAAGAAG CCGGTATTGT 7260
 TCTTACAGAT TTCCAAAAAG GAAATGAAAA AGCGCGTGAA CGTATGAAAG TACAATTTTC 7320
 AATTGCTTCA AACCGACAAG GTATTGTAGT AGGAACAGAT CATTGAGCTG AAAATATAAC 7380
 55

TAAACGACAA GGTGCTCAAT TATTAGCGTA TCTTGGTGCG CCAAAGGAAT TATATGAAAA 7500
 AACGCCAACT GCTGATTTAG AAGATGATAA ACCACAGCTT CCAGATGAAG ATGCATTAGG 7560
 5 TGTAACCTTAT GAGGCGATTG ATAATTATTT AGAAGGTAAG CCAGTTACGC CAGAAGAACA 7620
 AAAAGTAATT GAAAATCATT ATATACGAAA TGCACACAAA CGTGAACCTG CATATACAAG 7680
 10 ATACACGTGG CCAAATCCT AATTTAATTT TTTCTTCTAA CGTGTGACTT AAATTAAATA 7740
 TGAGTTAGAA TTAATAACAT TAAACCACAT TCAGCTAGAC TACTTCAGTG TATAAATTGA 7800
 AAGTGTATGA ACTAAAGTAA GTATGTTTCAT TTGAGAATAA ATTTTTATTT ATGACAAATT 7860
 15 CGCTATTTAT TTATGAGAGT TTTCGTACTA TATTATATTA ATATGCATTG ATTAAGGTGA 7920
 GGTGGAAGCA GTTTGGTATT TAAAGTGTA TGAAGAGA GTGGGGCGCC TTATGTCATT 7980
 CGTAACAGAA AATCCATGGT TAATGGTACT AACTATATTT ATCATTACG TTTGTTATGT 8040
 20 AACGTTTTTA ACGATGCGAA CAATTTTAAC GTTGAAAGGT TATCGTTATA TTGCTGCATC 8100
 AGTTAGTTTT TTAGAAGTAT TAGTTTATAT CGTTGGTTTA GGTGTTGTTA TGTCTAATTT 8160
 AGACCATATT CAAAATATTA TTGCCTACGC ATTTGGTTTT TCAATAGGTA TCATTGTTGG 8220
 25 TATGAAAATA GAAGAAAAAC TGGCATTAGG TTATACAGTT GTAAATGTAA CTTCAGCAGA 8280
 ATATGAGTTA GATTTACCGA ATGAACTTCG AAATTTAGGA TATGGCGTTA CGCACTATGC 8340
 30 TGCCTTTGGT AGAGATGGTA GTCGTATGGT GATGCAAAT TTAACACCAA GAAAATATGA 8400
 ACGTAAATTG ATGGATACGA TAAAAAATTT AGATCCGAAA GCATTTATCA TTGCGTATGA 8460
 ACCTCGAAAC ATACATGGTG GATTCTGGAC TAAAGGCATT CGTCGTAGAA AGCTTAAAGA 8520
 35 TTATGAACCA GAAGAACTGG AAaGTGTAGT AGAaCATGAA aTTCmAAGTA AaTGAGAAaTG 8580
 AAmCAATtGC TGATTGTTTG TCACGAATGA AatGCAAGGG TATATGCCGG TAAAACGTAT 8640
 TGAATAAACC GTGTTTCAAG AGCAAAAAGA TGGCACGGTT GAAGTATCAC ATCAAGAAAT 8700
 40 CGTTTTTGTA GGTAAGAAAA TCCAATAACA TAATCCAATT TAAATAAAGA CTATTTGAAG 8760
 AGGAAAGGCT ATTCAAAGTT TGAGTAATTT TACTTTGAAT AGCCTATTTG TTTATACATG 8820
 CAAGATGCTC GATCCATATT GTATGAGAAA CCCCCAGCAA GCTATATAAA GCATATGCTG 8880
 45 GGGGTTCTTA ATATTTTAAA AATTATTGTT AGATTATATA TATCGTCGCT TTTTCTAAAA 8940
 CAATCTCATC GCATGAAATT TTTTCTTCCT AGAGACCTTT AATAAGATTA ATAGTTTACT 9000
 50 TAATCATATC TAGATAGTCT TATGACTTAT GCTTAATGAA AGTCATTCTA GGAGAAGTTC 9060
 CCAAAGCTTC TGTGTTTATA ATTGTTAGTA GTATTTTATT ATCATTTGGT ATAAATATTT 9120
 CAATAACAAT TGAGCTATTA TTTTATTAT ATAATGTGAG TTGTTTGTGT TCTGTATTTA 9180

55

CATTTAAATC TTGAGGATGC CATTCTCCCT CAATAATATT AAGATAATAC TTAGCCTCTG 9300
 AATTACATTT GAATTTATCA ATACTAAATA ATTCAATTTG TTCCATAATA TTATTTACCT 9360
 5 TTCTAAAATA CAAATTTTAA TAACCATAAA TAGATGAATA CCATCGATAA TGGTCGCCAT 9420
 TGGATACTGG AATAACATTG TTTTGTAGCAT CTTGAGTCAT AAAACCATTG TCCCATGGAT 9480
 TCCATATAAT TATAACCTCT TGTCCATTAT CTAATTTAGC GTTCCCAACA ACTGCCATGG 9540
 10 CATGCCCTGC GTGCATACCA TTTCTTGATT CTACTCTACT ACCTAAAACA GCAATTCCTT 9600
 TATTATTTTT AGTAAGATTG TCAACTTCAT TATATGTAGT CATTCTATTA AGAAGTTGTG 9660
 GACTTCTTCC CTGAGTTTGT CCAAAATAAA TCATCTCTCT TGGCGTTAAA CCAGTAAATT 9720
 15 GGAATCGTTG TCCTTGTAAG TTTGGGTGTA AAAATCTCAT CACAGCTTCT GCATGATATT 9780
 TGTTAGTATT ATAAGTCGCA TTTAGTAATT CAGACATCGT ATAGCCTGCA CACCAACCAT 9840
 20 TGTTACCTTG AGTTTCTCTT ATCTTGAAAT TCTCAAGTTT ATTTATATAT TGSTCGTTGT 9900
 AAGTATAATT ATTACTTTTA AATTGACTAG TTGGCATAGT GACAGAAGCT TTTTGCTTTA 9960
 GTTGCCTTAC ATTATTGCCA GTAGGTATAC TCTCAGTCTT TnTnAACTnT nTATCTTCTA 10020
 25 GACGTGGTGT TTTTAGTACT AGTTTAGCTT TATGATTTTG AGTACCACAT AGTAACCTTT 10080
 TGAGTTGT 10088

(2) INFORMATION FOR SEQ ID NO: 33:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7563 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

40 CGGAAACGnA CCCnATGCGT ATGCTTGACG TGCCAAAATT AAATACGAAG TTCATAGCTT 60
 TGAGGTACCA GAAGAACATT TATCTGGTCA AGAAGTCGCA GnACTCATAC AAGCAAATGT 120
 TAAAACAGTA TTTAAAACGC TTGTCTAGA AAATACAAAA CATGAACATT TTGTATTTGT 180
 45 TATCCAGTA AGTGAAACTT TAGATATGAA AAAGGCAGCT GCTTTGGTTG GAGAGAAGAA 240
 ATTGCAGCTT ATGCCTTTAG ATAATTTGAA AAATGTAACG GGATACATTC GTGGTGGGTG 300
 TTCGCCTGTT GGTATGAAAA CATTGTTTCC AACAGTCGTT GACAAATCGT GTGAAAATTA 360
 50 TAGTCATATC AGTGTGAGTG GTGGGCTTCG AACAAATGCAA ATCACAATAG CTGTTGAGGA 420
 TTTGATTACA ATAACATAAG GCAAAATTGG AGCAGTTATC CATGAATGAT TAATAACAAC 480

55

	TGCCCACTC	CTTTTGTATT	GAATTAGCAT	TTTACGATCA	TAAACAGTCA	TTATAATTGA	600
	GTATTTGAAC	ATAAAAAATGT	AATTTTATCG	TAACAATTTG	AGTGTGTTGTG	ATTGTTTTTG	660
5	GTAATTTATG	ATTGAAAAGT	GAAAGCGTAC	TCATTATAAT	ACAAAGTGAG	ATGGGGTGAT	720
	GATGATAATT	ACTGAAAAAA	GACACGAGTT	AATATTAGAA	GAACTTTCGC	ACAAAGATTT	780
	TTTGACTTTA	CAAGAATTAA	TAGATCGAAC	TGGTTGCAGT	GCTTCAACAA	TACGArGAGA	840
10	TTTATCTAAA	CTACAACAAT	TAGGGAAATT	GCAACGTGTG	CATGGTGGTG	CAATGTTAAA	900
	AGAAAATCGT	ATGGTTGAGG	CGAATTTAAC	TGAAAAATTA	GCAACGAATC	TTGATGAAAA	960
	GAAAATGATT	GCTAAAATAG	CAGCTAATCA	AATCAACGAT	AATGAATGCT	TATTTATCGA	1020
15	TGCTGGTTCA	TCTACATTGG	AGCTAATTAA	ATATATTCAA	GCGAAAGATA	TCATTGTGGT	1080
	AACCAATGGT	TTAACACATG	TAGAAGCTTT	ACTTAAAAAA	GGTATTAAAA	CAATTATGCT	1140
20	AGGTGGTCAA	GTTAAAGAAA	ATACACTTGC	TACGATTGGT	TCTAGTGCTA	TGGAGATATT	1200
	AAGACGATAT	TGTTTCGATA	AAGCTTTTAT	CGGGATGAAT	GGATTAGATA	TTGAACTTGG	1260
	ATTAACACT	CCCGATGAGC	AAGAGGCATT	AGTTAAACAA	ACAGCAATGT	CATTAGCCAA	1320
25	TCAATCATTT	GTACTTATAG	ATCATTCTAA	GTTTAATAAA	GTATATTTTG	CTCGTGTAAC	1380
	TTTGCTAGAA	AGTACGACAA	TCATCACATC	TGAAAAAGCA	TTAAATCAAG	AATCGTTAAA	1440
	AGAATACCAA	CAAAAGTATC	ACTTTATAGG	AGGGACTTTA	TGATTTATAC	AGTGACTTTC	1500
30	AATCCTTCAA	TTGACTATGT	CATTTTTACG	AATGATTTTA	AAATTGATGG	TTTGAACAGA	1560
	GCAACAGCAA	CATATAAATT	CGCTGGGGGG	AAAGGTATTA	ATGTCTCGCG	CGTCTTAAAG	1620
35	ACATTGGATG	TTGAGTCAAC	TGCCTTGGGA	TTTGCAGGTG	GATTTCTCGG	GAAATTCATT	1680
	ATAGATACAT	TAAATAACAG	TGCAATTCAA	TCGAATTTTA	TTGAAGTTGA	TGAAGATACA	1740
	CGTATTAATG	TGAAATTAAA	AACAGGACAA	GAAACAGAAA	TCAATGCACC	GGGTCCTCAT	1800
40	ATAACGTCAA	CACAATTTGA	ACAACGTGTA	CAACAAATTA	AAAATACAAC	AAGCGAAGAT	1860
	ATAGTTATTG	TTGCTGGAAG	TGTACCAAGT	AGTATTCCAA	GCGATGCGTA	TGCGCAAATT	1920
	GCACAAATTA	CAGCACAGAC	AGGTGCTAAA	TTAGTAGTCG	ACGCTGAAAA	AGAATTGGCT	1980
45	GAAAGCGTTT	TACCATATCA	TCCACTATTT	ATTAAACCTA	ATAAAGATGA	ATTAGAAGTG	2040
	ATGTTTAATA	CAACAGTGAA	CTCAGACACA	GATGTTATTA	AATATGGTCG	TTTGTTAGTT	2100
	GATAAAGGTG	CGCAATCTGT	TATTGTCTCG	CTTGGCGGTG	ATGGTGCTAT	TTATATTGAT	2160
50	AAAGAAATCA	GATTAAAGC	AGTTAATCCA	CAAGGGAAAG	TGGTTAATAC	AGTTGGCTCT	2220
	GGTGATAGTA	CAGTTGCAGG	CATGGTGGCT	GGAATTGCTT	CAGGTTTAAC	GATTGAAAAA	2280
55							

	CGGGACGCTA TAGAAAAAAT AAAATCACAA GTTACGATTA GCGTACTTGA TGGGGAGTGA	2400
	AAATAATGAG AGTAACAGAG TTATTAACAA AAGATACAAT AGCAATGGAT TTAATGGCAA	2460
5	ATGACAAAAA TGGTGTATT GATGAGTTAG TAAATCAATT AGACAAAGCA GGTAAATTAA	2520
	GTGATGTCGC GTCATTTAAG GAAGCGATTC ACAATCGAGA ATCACAAAGT ACAACTGGTA	2580
	TCGGCGAAGG TATTGCCATT CCACATGCCA AAGTGGCCGC AGTTAAGTCA CCAGCTATTG	2640
10	CGTTTGGTAA ATCTAAAGCA GGCCTAGATT ATCAAAGTTT GGATATGCAA CCAGCACACT	2700
	TATTCCTTAT GATTGcAGcG CCAGAAGGTG GCGCCCAAAC ACATCTAGAT GCTTTAGCTA	2760
	AGTTGTCTGG TATTTTAATG GATGAAAATG TACGTGAGAA ATTATTACAT GCTTCATCAC	2820
15	CTGAAGAAGT ACTAGCGATC ATAGATGAGG CTGATGATGA AGTGACAAAA GAAGAAGAGG	2880
	CAGAAGCTGA AGCACAACAA GTTGCAACTG CAGAACAATC ATCTAAACAA TCTAATGAGC	2940
20	CATATGTGTT AGCAGTAACT GCTTGTCCAA CAGGTATTGC ACACACATAT ATGGCACGTG	3000
	ATGCATTGAA AAAGCAAGCG GATAAAATGG GTATTAAAT TAAAGTAGAA ACGAATGGTT	3060
	CAAGCGCAT TAAAAACCAT TTAAGTGAAC AAGATATTGA AAATGCAACA GGTATCATTG	3120
25	TTGCTGCTGA TGTTCATGTT GAGACGGATC GCTTCGATGG TAAAAATGTC GTAGAAGTAC	3180
	CAGTAGCAGA TGGTATTAAA CGCCCAGAAG AATTAATTAA TAAAGCATTG GATACAAGTC	3240
	GTAAACCTTT TGTGCCCCGT GATGGTCAAA GAAAAGGTAA CTCAAATGAC AGTCAAGAAA	3300
30	AATTAAGCCC AGGTAAAGCA TTCTATAAAC ACTTAATGAA CGGTGTCTTCT AACATGTTGC	3360
	CACTTGTAAT ATCTGGTGGT ATTTTAATGG CAATTGTATT TTTATTTGGA GCAAATTCAT	3420
	TTAATCCAAA AAGCTCAGAG TACAATGCGT TTGCAGAGCA GCTTTGGAAC ATTGGTAGTA	3480
35	AAAGTGCATT CGCGTTAATC ATTCCAATTT TATCTGGATT CATTGCACGT AGTATTGCGG	3540
	ATAAACCTGG TTTCGCTTCA GGTCTTGTAG GTGGTATGTT AGCAATTTCA GGTGGTTCAG	3600
40	GATTTATTGG TGGTATTATT GCAGGTTTCT TAGCAGGTAA CTTAACACAA GGTGTTAAAG	3660
	CCATGACACG TAAGTTACCA CAAGCATTAG AGGGATTAAA GCCAACATTA ATTTATCCAC	3720
	TATTAACAGT GACGGCTACA GGCTTATTGA TGATTTATGC CTTAATCCA CCAGCATCTT	3780
45	GGTTAAATCA TTTGTTATTA GATGGATTAA ACAATTTATC AGGTTCTAAT ATTGTATTAT	3840
	TAGGTTTAGT TATTGGCGCT ATGATGGCGA TTGATATGGG CGGTCCATTC AACAAAGCGG	3900
	CATATGTTTT TGCAACAGGT GCGTTGATTG AAGGTAATGC AGCACCAATT ACAGCTGCAA	3960
50	TGATTGGTGG TATGATTCCA CCGTTAGCAA TTGCGACAGC GATGTTAATT TTAGACGTA	4020
	AATTTACAAA AGAACAACGT GGTTCATTA TCCCTAACTA TGTGATGGGT ATGTCATTTA	4080

55

	TGATTGGTTC AGGTATAGGT GCGCAATTG CTTTAGGCTT AGG TTCACGA ATTACTGCGC	4200
	CACATGGTGG TATTATTGTA ATTGTTGGTA CTGATGGTGC AACTTACTT CAAACTCTTA	4260
5	TTGCACTTCT AGTTGGCACA TTAGTTTCAG CATTAAATTA CGGTTTAATC AAACCAAAGT	4320
	TAACTGAAAC AGAAATCGAA GCTTCAAAAT CAATGGACGA GTAGTTTTAA TGATGTAAAA	4380
10	TGATTGTTAG CAAAGAGCTT CATATTAAGT TGTATGTTCA ATGAATATAT GTTAGTTTTA	4440
	TATATCGTGT TAACGGTAGC TTATACAAAG CTGTAAAAAC ACTTCTATT AATTCAGTTT	4500
	TTATGAATTG ATATGAAAGT GTTTTTATTT TTAGATAAAT GAATGAAGAA ATAGACACCA	4560
15	CAAATGTATA GACTTTTTTA ATATTTTGCA AAAAGTTATG CCAAACGAAG CAGATATAGT	4620
	AAAATATGAG TGTCTTAAAG TGAAAATTA TAAATAAAGA AGGGTTTATA CGTGTCAGAA	4680
	TTAATTATAT ATAACGGCAA AGTTTATACT GAAGATGGCA AAATCGATAA TGGTTACATT	4740
20	CATGTGAAAG ATGGACAGAT TGTGCAATT GGAGAAGTGG ATGATAAAGC AGCAATTGAT	4800
	AATGATACGA CAAATAAAAT TCAAGTGATT GATGCTAAAG GTCATCATGT ATTACCAGGT	4860
	TTTATTGATA TACATATTCA TGGTGGTTAT GGTCAAGATG CAATGGATGG GTCATACGAT	4920
25	GGCTTAAAT ATCTATCCGA AAATTGTGTG TCTGAAGGGA CGACATCATA CTTGGCCACT	4980
	ACAATGACGC AATCGACTGA TAAAATAGAT AATGCACTTA CAAATATTGC TAAATATGAA	5040
30	GCGGAgCAAG ATGTTCACAA TGCAGCGGAA ATTGTAGGTA TACATTTAGA AGGACCATTT	5100
	ATATCTGAAA ATAAAGTTGG TGCTCAACAT CCGCAATACG TTGTACGCCC ATTTATCGAT	5160
	AAAATTAAAC ATTTTCAAGA GACTGCTAAC GGATTAATAA AGATTATGAC GTTTGCACCT	5220
35	GAAATTGAAG GTGCAAAAGA AGCGCTTGAA ACGTATAAAG ATGACATTAT TTTTCAATT	5280
	GGTCATACAG TAGCAACATA CGAAGAAGCA GTTGAAGCTG TTGAGCGAGG AGCTAAACAT	5340
	GTCA CGCATT TATATAATGC AGCGACGCCA TTCCAACATA GAGAACCAGG TGTTTTTGA	5400
40	GCAGCATGGT TGAATGATGC TCTACATACC GAAATGATTG TTGATGGCAC TCATTCTCAT	5460
	CCGGCATCGG TTGCAATTGC TTACCGTATG AAAGGTAATG AACGTTTTTA TTTAATTACC	5520
	GATGCAATGC GTGCAAAAGG TATGCCTGAA GGAGAATATG ATTTGGGTGG AAAAAAGTA	5580
45	ACTGTTCAAT CGCAACAAGC ACGTCTTGCA AATGGTGCGC TTGCTGGTAG TATTTTAAAA	5640
	ATGAATCATG GGTTACGTAA CTTAATATCA TTTACAGGTG ATACATTAGA TCATTATGG	5700
50	CGAGTAACAA GTTTAAATCA AGCCATTGCA TTAGGTATCG ATGATAGAAA AGGTAGTATT	5760
	AAAGTAAATA AGGATGCAGA TCTGTATTATT CTAGATGATG ATATGAATGT AAAATCTACA	5820
55	ATAAAACAAG GCAAGGTTCA CACATTTAGC TAATAAATAA TCATAATTAA ATGTATGCAA	5880

	TTTTCTGGGG GTGTCTAAAT GGGGAAGGCGA TAACATGTAG TTGTAATTTA AGTCATAGTG	6000
	ATAAATTTGA ATGCGTGTTA CCCATGAGTG ACACATATAA CATGGAGGTG AATCCCTAGA	6060
5	AATAGGGAAT TAATTGGAAG CTTCGACCAT AATTAGTTTG ATTATATTTA TTCTATTAAT	6120
	TGCATTAACC ACTGTATTTG TTGGTTCAGA ATTTGCATTA GTAAAAATTA GAGCAACAAG	6180
	AATTGAACAG CTAGCAGATG AAGGAAATAA ACCTGCTAAA ATAGTAAAAA AGATGATTGC	6240
10	TAATCTAGAT TATTATCTTT CTGCTTGTC A GTTAGGTATA ACAGTAACAT CTTTAGGGTT	6300
	AGGTTGGCTT GGTGAACCAA CGTTTGAAAA GCTATTACAC CCAATATTTG AAGCAATCAA	6360
	TTTACCAACT GCATTAAACG CGACGATTTT GTTTGCAGTG TCATTATATA TCGTTACGTA	6420
15	TTTGCATGTA GTACTTGGTG AATTAGCGCC TAAATCTATA GCTATTCAAC ATACTGAAAA	6480
	GCTTGCTTTA GTATATGCAA GACCATTGTT CTATTTCCGT AACATTATGA AACCATTGAT	6540
20	TTGGCTGATG AATGGTTCTG CACGTGTTAT TATTAGAATG TTTGGTGTA ATCCTGATGC	6600
	CCAACTGAT GCAATGTCAG AAGAAGAAAT CAAAATTATT ATTAACAATA GTTATAATGG	6660
	TGGAGAAATC AACCAAACCTG AATTGGCATA TATGCAAAAT ATCTTTTCAT TCGATGAAAG	6720
25	ACATGCAAAA GATATAATGG TACCTAGAAC TCAAATGATT AACTAAATG AACCTTTTAA	6780
	TGTAGACGAA TTACTAGAAA CAATAAAGA ACATCAATTT ACGCGTTATC CAATTACTGA	6840
	TGATGGTGAT AAAGACCACA TTAAAGGATT TATTAACGTC AAAGAATTTT TAACTGAATA	6900
30	CGCTTCTGGA AAAACGATTA AAATAGCAAA CTATATaCAT GAGTTGCCAA TGATTTTCAGA	6960
	GACAACACGT ATCAGTGATG CATTAATTAG AATGCAACGT GAACATGTAC ATATGAGTCT	7020
	TATTATAGAT GAATATGGTG GAACGGCAGG TATTTTAACG ATGGAAGATA TTTTAGAAGA	7080
35	AATCGTTGGA GAAATTCGTG ATGAATTTGA TGATGATGAA GTGAATGATA TCGTTAAAT	7140
	TGATATAAG ACATTCCAAG TAAATGGCAG AGTACTATTG GATGATTTAA CTGAAGAGTT	7200
40	CGGTATAGAA TTTGATGACT CTGAGGATAT TGATACGATA GGTGGATGGT TACAATCTCG	7260
	TAATACCAAT TTACAAAAAG ATGATTACGT GGATACAACT TATGATCGCT GGGTTGTTTC	7320
	AGAAATCGAT AACCACCAAA TTATTTGGGT GATATTAAAC TATGAATTTA ATGAAGCGAG	7380
45	ACCTACTATC GGACAGTCTG ATGAAGATGA AAAATCAGAA TAGATATTAA TATATAAACC	7440
	AACTAAGAAT GATTTAATTC ATTTTGGTT GGTTATTTTT TTGACTAAAA TTAAnGAAAA	7500
50	GTGAAAATAG TATTGGAAC CAATATCTTT AATGATTTAA TGAATAAnTT TTATTGAAAG	7560
	CGA	7563

(2) INFORMATION FOR SEQ ID NO: 34:

55

(A) LENGTH: 3492 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

10	TTATATCAAC TTCATGGCGG AACCATTGAT GACCCATTAG ACGAAACAAT AAGCGCATT	60
	ATGAATTGA AACAAGAAGG AATTATACGT GCTTACGGTA TTTCTTCTAT TCGCCCAAAT	120
	GTAATTGATT ATTATTTAAA ACATAGTCAA ATCGAAACGA TAATGTCTCA ATTCAATTTG	180
15	ATTGATAATC GTCCAGAATC ATTATTAGAT GCAATTCACA ACAATGATGT TAAAGTATTG	240
	GCAAGAGGAC CTGTGTCTAA AGGATTATTA ACTTCAAACA GTGTTAATGT GCTCGACAAT	300
	AAATTTAAAG ATGGTATTTT TGATTATTCT CATGATGAAT TGGGTGAAAC AATAGCCTCT	360
20	ATTAAAGAAA TTGAAAGTAA TTTATCTGCA TTGACATTTA GTTATTTAAC ATCACATGAC	420
	GTGCTTGGTT CCATCATTGT AGGTGCAAGT AGCGTCGACC AATTAAAGA AAATATTGAA	480
	AACTATCATA CTAAAGTTAG TTTAGATCAG ATTAAAACAG CAAGAGCTCG TGTAAGGAT	540
25	TTGGAATATA CCAATCATTT AGTGTAAG TCATTTTCAG TAATAAAAAC AGCAGCATGA	600
	GGCGTTTCAT TATAAAATG CCTTACTGCT GTTGTTTATG TACAATTCGC TATAATTTAT	660
30	GATTATGATT ACTCACTTAT GATAGAAAT AAAGCGTTGT CCTCACGCAT CAGTATTTAG	720
	TAATTTCGCC TTGCGGCATT GCCTTAAGCA AACTTCTGCC ACTTCATCTC TTAATAATTT	780
	TATTAACA TCTTTCTATA TTTCACTTCG CATGTTGATT CATCATTATT AGTTATTATT	840
35	TGTACACCCA GCACATTTCC TTGCAACACA AGTAGTTTGA ATTTTTCACA AGTATAATAT	900
	AATGTACCGT CTGAAATTTG GTCTACAGAA ATATCGCCTA AAATATCCAG CACTGTAAAT	960
	TCTTCAAATA CTGATAGTTG TTCCGCATAT CGTACACAAA GTCTTACCAC ACTCTCCGAT	1020
40	TGACAGTTCA TTGCCATCCC ACCTATTTAT GCTTTATTTT TAAATAATTT AGGGAAACAT	1080
	CGTTCAAAAA ATCTAGGCGC AATTTGATAC ATTTTCAACG CATGATGCAT CCATTTAGGC	1140
	CGATTAAATTT CCAATTGTTT TGTTTTAATG CCATAAATGA TATCTTCTGC AAGCTGATTA	1200
45	GCATCAAGCA TAATTTCCCC CATCTTTTTA GCATACTTCA TTGATGGGTC GGCTTTTGA	1260
	TGAAAAGGTG TATCAATCGG GCCAACATTA ACTGTCATGA TATGTAAGTT TGGTGACTCT	1320
50	AGTCTTAAAG CATTCAATTA TGCATAAAAC CCTGCTTTTCG ATGCCCCATA ATGTGCAGCA	1380
	TTTGCTTGTG TGGAAAATGC AGCTTGACTT GAAATACCTA CAATATGTGC GTTAGATGTT	1440
	AAATATGGTC TCAACACAGT ATATAAACA TTAAACTAA TTAAATTAAG CTGATACGTT	1500

55

	TAAATGAATC CATCGAATGA TGTATTGTCT TCAAATTGCA GTGCCTGTAT CGACTTCAAA	1620
	TCATTTAAGT CACAAGGAAT AACATTTATA GTTTTCCCCA ATTCCTGTTC AAAGATTCTA	1680
5	GTTGCTTTAT CAACATCACG CACCAACAAC GTTACATGCA CTTTATTTTC TAGTAACTTT	1740
	CGGACAATCG ATAAACCTAA ACCACTCGTA CCACCAGTCA CTATAAAATG TTGTCCTTTC	1800
10	ATCAATTAAC CTTCCCTTTTC AATTATATAG AATGCAATTT ATCAACTTTA CATAATTGAG	1860
	ACAAGTTGAT TATCTTTTCCT AATATATATA CAATAATAAG AAAATATAAC ATACAAATCA	1920
	AAAATAAAG GGATGTGACG TTAATGrAAC TCGTATTTTA TGGAGCTGGT AATATGGCAC	1980
15	AAGCTATATT TACAGGrATT ATTAACTCma GCAACTTAGA TGCCAATGAT ATATATTTAA	2040
	CAAATAAATC TAATGAACAA GCTTTAAAG CATTCGCTGA AAAACTAGGT GTTAACTATA	2100
	GTTATGAtGA TGCGACATTA TTAAGAGATG CAGAyTATGT ATTTTtagGT ACCAAACCAC	2160
20	ATGACTTTGA TGCTCTAGCA ACACGCATCA AACCACATAT TACAAAAGwC AATTGCTTCA	2220
	TTTCAATTAT GGCAGGTATT CCGATTGATT ATATTAAACA ACAATTAGAA TGCCAAAATC	2280
	CaGTTGCTAG AATTATGCCA AACACAAATG CGCAAGTTGG ACACTCTGTT ACTGGCATTa	2340
25	GTTTTTCAAA CAACTTTGAC CCTAAATCTA AAGATGAAAT TAACGATTTA GTTAAAGCAT	2400
	TTGGTTCTGT AATTGAAGTA TCAGAAGATC ATTTACATCA AGTAACAGCT ATCACC GGAA	2460
30	GCGGCCCAGC ATTTTTATAT CATGTATTCTG AGCAATATGT TAAAGCTGGT aCsAAACTTG	2520
	GTCTAGAAAA AGAACAAGTT GAAGAATCTA TACGCAACCT TATTATAGGT ACAAGTAAGA	2580
	TGATTGAACG TTCAGAtTTG AGCATGGCTC AATTAAGAAA AAATATTACC TCTAAAGGTG	2640
35	GTACGACACA AGCTGGCCTT GATACATTGT CACAATATGA TTTAGTATCT ATTTTCGAAG	2700
	ATTGTCTAAA CGCTGCCGTC GACCGTAGTA TTGAACTTTC TAATATAGAA GACCAATAAA	2760
	AACAaACCCG CCAACACATG TATGCATCAT CGCAAGCACT GTGTTTGACG GGTtATTTTT	2820
40	ATAATTTATT GTTATTTGGC AAGCATTGTT TATTACTTTG TCATTAGATT TTAAACTAT	2880
	CAAAATCTTT TACAAAATTA AAATTAGGTG TATCTTCATT TTGTATCAAT GTTTGATAAA	2940
45	TTTCATTTAT ATCTTCTGTA TTATAGCGAT TGCTCAAATG TGTAATCAAC GTACGTTTAA	3000
	CATTGGCTTC TTTTATCAAT GCAAATACGT CTTCAATATG GCTATGATGA TAATTGTTGG	3060
	CTAAATGCTT TTCACCATCT ATATAGGTCG CTTCATGTAC CATCACATCA GCATCTCTAG	3120
50	AAATCACACG TTCATTAGAA CATGGTTTTG TATCACCAAA AATTGCTACA ACTGGACCCT	3180
	GTTTGGACTC ACCTCTAAAA TCTTTTGATT GATAAACTTG ACCATTATGT TCAAATGTAT	3240
	CATGAGATTt TACTTCTTGA TATTTAGGAC CTGGTTCAAG ACCAATGTTT TTTAACGCTT	3300

55

CATGATTAAG TAAATGCGCC TCTACAGTAA AACCATCCAT GATGATATGT CAGATGATCA 3420
 TCGATTTCAA TATATGtAAT TGGATAGTTT AAATGTGACT CTGATAAATT CATAGACATT 3480
 5 TCCACATATG CT 3492

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1973 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

ATCTAGCGGT ACAAGCGTCT TGGAGGCTAG TATGTTGAAC ATTGTAAACC CTGAAGATCA 60
 20 CTTGCGTTGTC ATTGTTTCAG GTGCCTTTGG TAACCGATTT AAACAAATTG CACAACTTA 120
 TTACAAAAAT GTGCATATTT ATGACGTAAC ATGGGGAGAA GCTGTAGATG TCAAAGATTT 180
 CATCAATTTT CTTTCAACTT TAAATGTTGA AGTTAAAGCA GTATTTAGTC AATATTGCGA 240
 25 AACATCTACG ACAGTGCTAC ACCCTATTCA CGAGTTAGGA AATGCCATTA ATCAATTTAA 300
 TAGTAATATT TATTTTGTAG TTGACGGCGT AAGTtGCATT GGTGCTGTTG ATGTTGACAT 360
 TAACAAAGAT AAAATTGATG TACTTGTTTC TGGTAGTCAA AAAGCAATTA TGTTACCTCC 420
 30 AGGATTAGCT TTTGTAGCTT ATAGCCACCG TGCAAAAGAA CATTTCAAAG AAGTAACTAC 480
 GCCAAATTT TATCTAGACT TAAATAAATA CATTTCGTCA CAAGCTGACA ATTCTACACC 540
 GTTCACACCA AATGTGTCTT TATTTAGAGG TGTAATGCA TACGTTGAAA CCGTAAAAGC 600
 35 AGAAGGTTTC AATCACGTAA TAGCACGACA CTATGCAATT AGAAATGCAT TAAGAAGCGC 660
 CTTAAAGCA TTAGATTTAA CTTTATTAGT CAATGATAAA GATGCATCTC CAACGGTTAC 720
 40 AGCATTCAAA CCTAATACAA ATGATGAAGT GAAAATAATC mAAGATGAAC TTAAAAATnG 780
 CTTTAAAATA ACAATTGcng GTGGTCAAGG CCATCTTAAA GGTCAAATTT TnAGAATTGG 840
 TCATATGGGG AAAATTAGTC CTTTCGATAT TTTATCGGTA GTATCTGCTT TAGAAATTAT 900
 45 TTTAACTGAA CACCGTAAAG TTAATATAT CGGTAAAGGT ATATCAAAAT ATATGGAGGT 960
 TATTCATGAA GCAATTTAAT GTACTCGTTG CAGATCCCAT ATCAAAAGAT GGTATCAAAG 1020
 CATTATTAGA TCACGAACAA TTCAATGTAG ATATTCAAAC TGGCTTGTC GAAGAAGCAT 1080
 50 TAATCAAAAT TATACCTTCA TACCATGCTT TAATCGTTTC TAGTCAAAC ACGGTTACTG 1140
 AAAATATCAT AAATGCTGCT GATTCTTTAA AAGTAATCGC ACGCGCCGGT GTTGGTGTAG 1200

5 GTAATACGAT TTCAGCTACT GAACATACAC TGGCAATGTT ATTATCAATG GCACGAAATA 1320
 TTCCGCAAGC ACACCAATCA CTTACAAATA AAGAATGGAA TCGAAATGCA TTTAAAGGTA 1380
 CTGAGCTTTA TCATAAAACA TTAGGTGTCA TTGGTGCTGG TAGAATTGGT TTAGGTGTTG 1440
 CTAAACGTGC GCAAAGTTTC GGAATGAAAA TACTAGCTTT TGACCCTTAC TTAACGGATG 1500
 10 AAAAAAGCAA ATCTTTAAGC ATTACGAAGG CAACAGTTGA TGAGATTGCC CAACATTCTG 1560
 ATTTGCTTAC ATTACATACA CCACTAACAC CTAAAACAAA AGGCTTAATT AATGCTGTCT 1620
 TTTTGGCCAA AGCAAAACCT AGTTTGCAAA TAATCAATGT GGCACGTGGT GGTATTATTG 1680
 15 ATGAAAAGGC GCTAATAAAA GCATTAGACG AAGGACAAAT TAGTCGGGCA GCTATCGATG 1740
 TGTITGAACA TGAACCTGCA ACTGACTCGC CTCTTGTTGC ACATGATAAA ATTATTGTTA 1800
 CACCTCATTT GGGTGCTTCA ACAGTCGAAG CTCAAGAAAA AGTGGCAATT TCTGTTTCAA 1860
 20 ATGAAATCAT CGAAATTTTA ATTGATGGTA CTGTAACGCA TGCAgTGAAT GCACCTAAAA 1920
 TGGACTTAAG CAATATAGAT GATACTGTAA AATCATTCAT CAATTTAAGC CAA 1973

(2) INFORMATION FOR SEQ ID NO: 36:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7620 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

35 GGTGTTTCAG ATGTCACTGG TTGATTTTAA ATTGTAGACG GGTATTTTGG GCTTTCGCCA 60
 TATTTATTTG CCGGCTTACT GTCAAAGCAT AGGAATACTA TCATAACAAT TGTTAGGCCT 120
 AAATGAACAA AATAAAGAAG TACTAACAAA ATATTAAGAC CCATCGGCAT TAATGTAAAA 180
 40 TCACTGTCAT AATAACTATC GATAATCTGT AATACTATAT AAAATATAAT ACTGAATACT 240
 GTCATAATCA TTGGAAATAA CATTGTTCTT GATATATCGT GAAATCTTCG AACGCACAAC 300
 GCTAAATTTG GAATAAACGT TGCCAAACTA TAGACAAAAG TATACACAGA TGTAAGGATA 360
 45 ATCATCAATA TACTCATAAC TATTAATGTT TCGTTATCCG CCGCTATAGA AATAAAGAAT 420
 AGAAATAGGT TTATTATTAG CACACACACA GCTGGAACCA TAAGTATCAA ATGCCATAGT 480
 GCCATATACC AATATTCACT ACGTCTTGAT CTCCCCTTAA AATTTACATA ATTTTCCAA 540
 50 AATAAACGA ATGATTTCAT AAAACCTACT TGAGGTAATT GTTCCATTGT AATCTCCCTT 600
 TCGTTAATCA TATTTATATT TTTAATTATT GTTACCGTTA TAATTTACAA GATTCATTAT 660

55

	GTAAATGAA AACCCGCTAC AAGTACACAT CTATATGGAG ACTCATTGGA AAGTCAACGC	780
	TTCGTAACT ATACTAAAA TATGTCATAC TGCAATGTTT ACGTTTAAAA GAGTCTCAAT	840
5	CTATGCAAAT AAAATATTCC ATAACAAAGT ATATACTTTA CATTTTTATA ATTCTTAACA	900
	ATACTATTTT ATCAAACATT TACCACAATA AAAATATCTT TTTTATTTT ATTAAATTA	960
10	ATCATATAAT TGCGAGGAGA ATATTATGGA TTTCGTAAAT AATGATACAA GACAAATTGC	1020
	TAAAACTTA TTAGGTGTCA AAGTGATTTA TCAGGATACC ACTCAAACGT ATACAGGCTA	1080
	CATCGTGGA ACGGAAGCTT ACTTAGGTTT GAATGATCGT GCGGCTCATG GCTATGGCGG	1140
15	TAAAAATAACA CCTAAAGTCA CGTCATTATA TAAACGTGGT GGTACAATT ATGCACATGT	1200
	CATGCATACG CATTACTCA TTAATTTTGT AACAAATCT GAAGGTATAC CTGAAGGCGT	1260
	ACTTATCCGC GCAATTGAAC CAGAAGAAGG TTTATCCGCT ATGTTCCGTA ACAGAGGTAA	1320
20	GAAAGGCTAC GAGGTAACGA ATGGCCCAGG AAAATGGACT AAGGCATTTA ACATTCCACG	1380
	GGCTATCGAT GCGCTACGT TAAATGACTG TAGATTGTCT ATTGATACTA AGAATCGTAA	1440
	ATATCCTAAA GATATTATTG CTAGTCCACG AATCGGTATT CCAAATAAAG GTGATTGGAC	1500
25	ACATAAATCT TTACGTTACA CAGTGAAAGG TAATCCATTT GTGTCTCGCA TCGTAAATC	1560
	AGATTGTATG TTTCCGAAG ATACTTGGA ATAAATGCCA TCTTTCATTG ATTACTATCA	1620
	TGAAAATGAA ATCTATCTCC TTATAAGTCA ATCAATCGTG CCGTCAACAT GCGGATGGGT	1680
30	TGATTGTTTT TCTTTGTATC CATCATATTT TTTGATTCAT CTCCTCTTAT TGAAC TTGTT	1740
	CTTAATTATA AAATATAACA ATAGAATTAT TTATAATTAT TAAATTTAGA TGCATTAATA	1800
35	TTATTGATAT TATTTTCAA AACTAGAAAT ATTGATTTGT TGCATGTATA ATGTTAAAAG	1860
	CGCCCTTTTA TAACGCTTAC ATATAAAAGC TTATTTAGGG AGAGGGATAT TCAACAAGGG	1920
	GGATTGAAA ATGATAGAAC TTAATGCAAT TACAACATTA TGTTTAGCTT GTATCCTTTA	1980
40	TTTACTTGGT AAGGCTATCG TTAATCACGT TAATTTTTTA AAACGTATTT GTATACCAGC	2040
	ACCAGTGATT GGCGGCTTAA TCTTTGCTAT TTTAGTTGCG GCTTTGGATT CATTGGCAT	2100
	GGTTAAGATT AAATTAGATG CTTCAATCAT TCAAGATTTC TTCATGTTAG CATTCTTTAC	2160
45	GACAATCGGT CTTGGTGCAT CATTGAAATT ATTTAAATTA GGTGGCAAAG TCTTGCTATT	2220
	ATACTTTATG TTTTGTGCTA TCATTTAGT CATTCAAAC ATAGTTGGTG TATCACTAGC	2280
	AAAAGTATTA AATATTAAAC CTTTGTAGG ATTAACAGCA GGTTCATGT CTATGGAAGG	2340
50	CGGTCATGGT AATGCTGCTG CTTATGGTAA GACAATTCAA GATTTAGGTA TTGATTCGGC	2400
	ACTGACAGCG GCTCTTGCA CTGCAACTTT AGGTCTTGTA TTTGGAGGGC TTATCGGTGG	2460

55

	ATTTAAAGAT TATAGCCAAG TAGCATATAA CGAACATTTA CATAGTAAAT TTAATGCCAC	2580
5	TGAAGTATTC TTCATTCAAT TTACAATCGT TGTATTCTGT ATGGCAGTTG GAAGTTATTT	2640
	CAGTCATTTG TTTACAGCTC AAACAGGGAT TAATGTTCCA ATTTACGTTG GCTCATTATT	2700
	TGTAGCTGTT ATTGTCCGAA ATATCTCTGA AAGTTTAAAT TTAAATATTG TAGATTTAAA	2760
10	AATTACTAAT CAAATTGGCG ATGTCGCATT AGGTATTTTC TTATCTCTTG CGCTAATGAG	2820
	CATTCAATTA ATCGAAATTT ATAACTTGC TATACCTCTT ATTATTATCG TTTTAGTTCA	2880
	AGTTGTCTGTT ATGATTTTAT TTGCTGTTTT AATTTTATTT AGAGGTTTAG GAAAAGATTA	2940
15	TGATGCTGCA GTAATGGTAG GTGGTTTTAT CGGTCATGGG CTTGGTGCAc GCCAAATGCC	3000
	ATGGCAAATT TAGATGTTAT TACTAAAAAA TATGGAACT CACCTAAAGC ATATTTAGTT	3060
	GTACCTATTG TTGGTGCATT CTTAATCGAT TTAATTGGTG TTATAGTCAT TATGGGATTC	3120
20	ATACAATGGT TTAGTTAAAC ACCAACTCA TAAATAAAAG AGGAGGCCTT CGCCTCcTcT	3180
	TTTATTTATC CTCGATGTAT ATTCAAGTTA CGTTGTTCTA TCCATGACAA TATTTCCGGA	3240
	CTAAATACGA TTTGTTTTTG TGTTAAGTCG TCAATATTTT TAGCATCTAA CATCGTCATT	3300
25	ATTGATTTCA TGTGTTCAAT AAATGATTCT ACATAAGCTA CTGTATGTGC AATGCCATTA	3360
	TTTTCAACTT GATTTAAAAA CGGACGTGAC ATACCAGTTG CCTTTGCACC AAGTGCTAAA	3420
30	CTTTTAATTG CATCGAGTGG TGTACGTAAA CCACCACTCG CGAAAACTGA AATTTGCTT	3480
	TGATAAGCCG TTGTTTCAAG TAATGACTCA ACTGTAGACT GTCCCATGA TGATAAGTAA	3540
	TCCATATCTT TATTTGCACG ACGTTCATTT TCAATATCTA CAAAGTTAGT ACCACCTTTG	3600
35	CCACTAACAT CGACATACTT GACGCCTATT TGTTGTAAGT CATGCATTAA TTCTTTGCTC	3660
	ATACCAAATC CAACTTCTTT TATAATGACT GGAACAGACA CTCGTGATAC AATCGACGCT	3720
	ATATTATCTA ACCAAGTCAC AAATTCACGA TTCCCTTCAG GCATAACTAA TTCTTGAGGA	3780
40	GAATTAACAT GGATTTGTAA CGCTTGTCGC TCAAGTAATT CAACTGCTTC CAAAGCCTTT	3840
	TCTACTGGTA CGTCCGCACC AACATTGCTA AAAATCATGC CTTCAGGATT CATTTTTCGC	3900
	GCAATCGTAA ACGTCTCAGC CATGCGTGGT TTTCTCAATG CCGCATGTGT TGATCCAACT	3960
45	GCCATCGCTA AGCCAGTTTC TCTTGCAACT ACAGCTAGCT TTTCATTGAT GTTTTTCGTC	4020
	CACTCGCTAC CACCCGTCAT TGCATTAATA TAAACCGGAT ATGCCATCGT TAAGTCAGGC	4080
50	GTCTGTGATG TCAAATCGAT ATCATTTACA TTAATTGATG GGATAGAATG ATGCACAAAA	4140
	CGCATCTTAT CAAAATCTGA ATGCATTGCG TCAGATTGGG CCATTGCTAT TTCAACATGT	4200
	TCATTTTTTC TCTGTTCTCT TTGAAAATCA CTCATGATTA AACCTACCTT TTCGTCATTT	4260

55

	ATTACAGCTA AGCAAATATA ATATCCATAA TGTAATGTA ATGCCGGCAT ATTTACAAAG	4380
5	TTCATACCAT AAATCCCAGC TATGAATGTT AACGGTGAAA ATATAACTGA TACTAATGTC	4440
	AGTACTTGCA TAATACTATT CATTCTAAAT GACGTGTATG ACTCAAAATT TTCTCGTATT	4500
	TCGTTTGTCA TTTCTTGAGC AGTACGAATG ATATTACGTT GCTTAATCAA GTGGTCATCG	4560
10	ATATGTTGAA TGTATAGCGA ATGTTTATTA TCTATAATCA AATCACCATT TTGTTTCATT	4620
	GSTATCAATTA GCTCTTGCAAT AGGAAACAGT ACACGTTTTA CTTTAATCAA ATCCGAACGT	4680
	AACTTAAAGA CACTATCCAT GACCATTITA TTAAAGCGAT CATCTACATG GCGGTCTTCA	4740
15	AAATGATAAA CACTATCTTC AAGTGCATAT ACAAAGTTGA AATATTTATC AACCATCATA	4800
	TCTAAAATTA ATATGACGAC ATCTGCACAA TCTAATTCTG CATCTAATGT ATTCATATAC	4860
	TTATAGACTA CTTTATTTAA TGATTCCAAC GTTTGATGAT GATATGTTAC TAATACATTG	4920
20	TCTTGTATAA AAATATTTAG TGCTATTGGT GAATAGTTTG ACCCCATAAT ACTATGGAAT	4980
	ACTAAGTATT GATAATCTTT ATAAGATTTA TATTTAGCTC GTGGCATACC GTTAATTGCA	5040
25	TCATCCACTT CTAAATCATT AAAATTAAAA TGTGCTTTAA ACCATTTCATT TTCTTGTTCA	5100
	TTCGGTTCAT CAAAATCATA CCAAACAATA GTCGCATCTT TTGGTATCTC TTTGATATCA	5160
	TCAACTACTT TAAACGGTTC ATATGTAGTT TGATACCGTA TCTTTAAAGC CATCGATACT	5220
30	CCCCCTAAAT AACGAATTCT CTATTATTTT ATCATGAATT AAATAACGTG TATGTCTTAA	5280
	TTTATTTTAG TATGATAGTC ACTAAGGAGA TGTTTATTAT CAAACAACCT TTTACACATA	5340
	CTCAAACCGT AACATCTGAA TTCATTGACC ATAACAATCA TATGCATGAT GCAAATTATA	5400
35	ATATCATTTT TAGTGACGTC GTGAATCGTT TTAATTACAG CCACGGTCTT TCTTTAAAG	5460
	AACGCGAAAA TTTAGCATAT ACGCTATTTA CACTAGAAGA ACATACGACA TACCTCTCAG	5520
	AATTGTCTCT TGGCGATGTA TTTACTGTTA CTTTATATAT TTATGATTAC GATTATAAGC	5580
40	GGTTGCATTT ATTTTAAACA TTAATAAAG AAGATGGTAC ACTAGCATCA ACAAATGAAG	5640
	TAATGATGAT GGAATTAAT CAGCACACAC GTCGTTCTGA TGCTTTTCCT GAATCATTTT	5700
	CAACACAAAT AGCACACTAT TATAAAAATC AATCAACTAT CACTTGGCCT GAACAATTAG	5760
45	GACATAAAAT AGCAATTCCA CACAAAGGAG CATTAAAATG ACAGATGCAT TACAACAAAA	5820
	GATTCATATC GAATTACTAG ATTTATTAGA TGATGTTAAG TTTGAATTAA CAGAATTAAA	5880
50	TGCACAAAAA GGGTTATACA TTAACGGACC AGCAAATCAG CTACTTAAGC GTGGCGTGCA	5940
	TATGGCTTAT GTTCAAGGAC AAAAGCAAGC CATCGATAAT ATTATGACTA TTGTGGAACA	6000
55	ACAGCTTGAA AGATCAACAT TTCCTAGAAC ATTATGATAA ATTTCAAAAT GAGGTTGCTC	6060

5 ATAATTTTTT AGATCAATTT TATCAAATTA AAGGGCAATA CTTTATCATC ACACATATCA 6180
 ATACACTTAT TGGTGATTTT CACTCAGAAG CTCATTAACA ATTAGTCTAT ATAACCCTTG 6240
 10 CTATATTTTC AAAAACAAAA CCCAATTACG TTTTCATGTC AAATATCATC TTGCATGAAA 6300
 TCGTAACTGG GTCATTTATA TGTTATTAGT TATTTTGTGT TACATCCTCA TCTATCGATT 6360
 TGGCAATTTG TTTAATAGCT TTATGTGATT GTCTAATTGG ATAAATTGGA AAATCATGTA 6420
 15 CCATCTTAGG ATAATCATAA AACTCAATGT ATTGATGATG TTGCAACATC ATTTGTTCAA 6480
 ATAGCTTCAT ATCAGGATGT GTCATTTTAC GTCCACCACC AAACATATAA ACTGGTGGCA 6540
 ATCCTTCTAT TGTGCCATTA ATTGGCGATA TGCCTTATC TGTTAATGGT AGGCCATTCTG 6600
 CCCATTTTTT CATAATCTCA TTGACACCAA ACTGACTTAG aACCGCATCT GTTTCGATTA 6660
 AGGCGTCCGA AATATCTTTA TTAGATAGTG TTGCATCTAA AATTGGTGAG ATTAAATACA 6720
 20 ATTTATTCGG TAATGGCTGT TGATTAKCTA AAAGAGATTG TACAAAGGAT AATGCCAGTG 6780
 CACCACCTGA ACCATCACCC ATGACTACGA CATTTTGATG TCCTACTTCA GATACTAATT 6840
 GaTCATAAAC ACGTTGTATC GCTTGGA AAA GTATCGTcaA TATGnaAACT CTGGTGTCTT 6900
 25 TGGATAGATA GGCAGTACAA CCTCATATAA TGtACTTAAA GTGATTTTAT CCCAACAATC 6960
 TCCAATGGAA CGGTGATGGT TGTAGTGCAT TGAATCCACC GTGAATATAT AAAATTTTCT 7020
 TATCAATTTG ATGTCTGAAA TTAAAGCGAA AGACTTGAT ATCATCTAAT GACAATTTTT 7080
 30 CTAAATTTGC TTTAACAATT AATGTTGAAG GCTGCTTATG TTTTTTCTA TTTTCAATTT 7140
 CTCTTTTATA AAAAAATCTT TCAACATCTT GATCATTTTT AAACATAATC GAGCGATTGT 7200
 GAAGCAAATA TTTATTGACA ACGCTATTCA TAACACGGTT TCTAATCAAT GTCTTAACCT 7260
 35 ACCTTTATAT ATTTTATGTA TCCAATGATk GTCTATCCCC TACATTCTTT GCCAAAAAAA 7320
 GTATATAATG TAGAAGATAT TTTCTTTTTC ACTTTCAAAT TTAAGACTAC AATTGAACAG 7380
 40 TGATTTTTCa TCATTATAAC AGACAAC TAGAATATTGAT AAGTAAAGAA AAGAACTTTA 7440
 TACGGAGGTA CCTTGCATGA CAAATCCAAA TCAACGATTA GAACCATTTG ATGAGACATT 7500
 TCAACAACCG AATATTCATC GTGGTAAGCG ATATGGTAAG AAAAAACGTT CATTGGTAAG 7560
 45 CATGATTATT CAAATCATTG TTGTWATATT AACCACCATC GCTGGAATAC AGCATGGTGG 7620

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 9834 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

	GTCATtACCG amTTTtTAG AaTCATTTAA AGATGATAAA TATACAAACG TTGGTAATTT	60
5	AAAAGAAGTG AATTTTGATA AAATTGCTGC GACGAAACCC GAAGTAATCT TTATCTCTGG	120
	ACGTACAGCT AATCAAAAGA ATTTAGATGA ATTCAAAAA GCTGCACCTA AAGCGAAAAT	180
	TGTTTATGTT GGTGCAGATG AAAAGAACTT AATTGGTTCA ATGAAACAAA AACTGAAAA	240
10	TATCGGAAAA ATTTACGATA AAGAAGATAA AGCTAAAGAA TTAAATAAAG ATTTAGATAA	300
	CAAAATTGCT TCAATGAAAG ATAAAACGAA AAACCTTCAAT AAAACTGTTA TGTATTTACT	360
	AGTTAACGAA GGTGAATTAT CAACATTTGG ACCTAAAGGT CGTTTTGGTG GATTAGTTTA	420
15	CGATACATTA GGATTCAATG CAGTTGATAA AAAAGTAAGT AATAGCAATC ATGGACAAAA	480
	TGTTTCTAAC GAATATGTTA ATAAAGAAAA TCCAGATGTT ATTTTAGCGA TGGATAGAGG	540
20	TCAAGCGATA AGTGGTAAAT CAACTGCGAA ACAAGCATTa AATAATCCTG TATTAAAAA	600
	TGTTAAAGCA ATTAAAGAAG ACAAAGTATA TAATTTAGAT CCTAAATTAT GGTACTTTGC	660
	AGCTGGATCA ACTACAATA CAATTAAACA AATTGAGGAA CTTGATAAAG TTGTAAAATA	720
25	ATTTTAAAAG AGGGGAACAA TGGTTAAAGG TCTTAATCAT TGCTCCCCTC TTTTCTTTAA	780
	AAAAGGAAAT CTGGGACGTC AATCAATGTC CTAGACTCTA AAATGTTCTG TTGTCAGTCG	840
	TTGGTTGAAT GAACATGTAC TTGTAACAAG TTCATTTCAA TACTAGTGGG CTCCAAACAT	900
30	AGAGAAATTT GATTTTCAAT TTCTACTGAC AATGCAAGTT GGCGGGGCCC AAACATAGAG	960
	AATTTCAAAA AGGAATTCTA CAGAAGTGGT GCTTTATCAT GTCTGACCCA CTCCCTATAA	1020
35	TGTTTTGACT ATGTTGTTTA AATTTCAAAA TAAATATGAT AGTGATATTT ACAGCGATTG	1080
	TTAAACCGAG ATTGGCAATT TGGACAACGC TCTACCATCA TATATTCATT GATTGTTAAT	1140
	TCGTGTTTGC ATACACCGCA TAAGATTGCT TTTTCGTTAA ATGAAGGCTC AGACCAACGC	1200
40	TTAATGGCGT GCTTTTCAAA CTCATTATGG CACTTATAGC ATGGATAGTA TTTATTACAA	1260
	CATTTAAATT TAATAGCAAT AATATCTTCT TCGGTAAAAT AATGGCGACA scgTGTTTCA	1320
	GTATCGATTA ATGAACCATA AACTTTAGGC ATAGACAAAG CTCCTTAACT TACGATTCCT	1380
45	TTGGATGTTT ACCAATAATG CGAACTTCAC GATTTAATTC AATGCCAAAT TTTTCTTTGA	1440
	CGGTCTTTTG TACATAATGA ATAAGGTTTT CATAATCTGT AGCAGTTCCA TTGTCTACAT	1500
	TTACCATAAA ACCAGCGTGT TTGGTTGAAA CTTCAACGCC GCCAATACGG TGACCTTGCA	1560
50	AATTAGAATC TTGTATCAAT TTACCTGCAA AATGACCAGG CGGTCTTTGG AATACACTAC	1620
	CACATGAAGG ATACTCTAAA GGTGTTTtag ATTCTCTACG TTCTGTAAA TCATCCATTT	1680

55

	AGTGTTCCTTT TTGAATAATG CTATTACGAT AATCTAACTC TAATTCCTTT GTTGTAAGTT	1800
	TAATTAACGA GCCTTGTTTCG TTTACGCAAA GCGCATAGTC TATACAATCT TTAACCTCGC	1860
5	CACCATAAGC GCCAGCATTG ATATACACTG CACCACCAAT TGAACCTGGA ATACCACATG	1920
	CAAATTCAGG GCCAGTAAGT GCGTAATCAC GAGCAACACG TGAGACATCA ATAATTGCAG	1980
	CGCCGCTACC GGCTATTATC GCATCATCAG ATACTTCGAT ATGATCTAGT GATAATAAAC	2040
10	TAATTACAAT ACCGCGAATA CCACCTTCAC GGATAATAAT ATTTGAGCCA TTTCCTAAAT	2100
	ATGTAACAGG AATCTCATTG TGaTAGGCAT ATTTAACAAC TGCTTGTAAT TCTTCATTTT	2160
	TAGTAGGGGT AATGTAAAAG TCGGCATTAC CACCTGTTTT AGTATAAGTG TATCGTTTTA	2220
15	AAGGTTCAAT AACTTTAATT TTTTCATTTG GGATAAGTTG TTGTAAAGCT TGATAGATGT	2280
	CTTTATTTAT CACTTCTCAG TACATCCTTT CTCATGTCTT TAATATCATA TAGTATTATA	2340
20	CCAATTTTAA AATTCATTTG CGAAAATTGA AAAGAAAGTA TTAGAATTAG TATAATTATA	2400
	AAATACGGCA TTATTGTCGT TATAAGTATT TTTTACATAG TTTTCAAAG TATTGTTGCT	2460
	TTTGCATCTC ATATTGTCTA ATTGTTAAGC TATGTTGCAA TATTGGTGT TTTTTGTAT	2520
25	TGAATTGCAA AGCAATATCA TCATTAGTTG ATAAGAGGTA ATCAAGTGCA AGATAAGATT	2580
	CAAATGTTTG GGTATTCATT TGAATGATAT GTAGACGCAC CTGTTGTTTT AGTTCATGAA	2640
	AATTGTTAAA CTTCGCCATC ATAACCTTCT TAGTATATTT ATGATGCAAA CGATAAAACC	2700
30	CTACATAATT TAAGCGTTTT TCATCTAAGG ATGTAATATC ATGCAAATTT TCTACACCTA	2760
	CTAAATATC TAAAATTGGC TCTGTTGAAT ATTTAAAATG aTGctACCGC CAATATGTTT	2820
	TGTATATTTT ACTGGGCTGT CTAAGAGGTT GAATAATAAT GATTCAATTT CAGTGTATTG	2880
35	TGATTGAAAA CAATTAGTTA AATCACTATT AATGAATGGT TGAACATTTG AATACATGAT	2940
	AAAGTcCTTT GATATTGAAA ATTAATTTAA TCACGATAAA GTCTGGAATA CTATAACATA	3000
40	ATTCATTTTC ATAATAAACA TGTTTTTGTA TAATGAATCT GTTAAGGAGT GCAATCATGA	3060
	AAAAAATTGT TATTATCGCT GTTTTAGCGA TTTTATTTGT AGTAATAAGT GCTTGTGGTA	3120
	ATAAAGAAAA AGAGGCACAA CATCAATTTA CTAAGCAATT TAAAGATGTT GAGCAAAAAC	3180
45	AAAAAGAATT ACAACATGTC ATGGATAATA TACATTTGAA AGAAATTGAT CATCTAAGTA	3240
	AACTGATAC AACTGATAAA AATAGTAAAG AATTTAAGGC ACTACAAGAA GATGTTAAAA	3300
	ACCATCTCAT ACCTAAATTT GAAGCATATT ATAAGTCAGC AAAAAATTTG CCTGATGATA	3360
50	CAATGAAAGT TAAGAAATTA AAAAAAGAAT ATATGACGCT TGCAAATGAG AAGAAGGATG	3420
	CGATATATCA ATTAAAAAAA TTCATAGGTT TATGTAATCA ATCTATCAAG TATAACGAAG	3480
55		

AATTAGCTGA TAATAAAAGT GAAGCAACTA ATCTTACGAC AAAATTAGAA CATAATAATA 3600
 AAGCGTTAAG AGATACTGCG AAGAAGAACC TAGATGATAG TAAAGAAAAT GAAGTAAAAG 3660
 5 GCGCGATTAA AAATCACATT ATGCCAATGA TTGAAAAGCA AATTACCGAT ATTAACCAAA 3720
 CTAATATTAG TGATAAGCAT GTTAATAATG CAAGGAAAAA CGCAATAGAA ATGTATTACA 3780
 GTCTGCAGAA CTATTATAAT ACACGTATTG AAACAATAAA GGTTAGTGAG AAGTTATCm 3840
 10 AAGTCGATGT AGATAAGTTG CCGAAAAAGG GTATAGATAT AACTCACGGC GATAAAGCCT 3900
 TTGAAAAAAA GCTTGAAAAA TTAGAAGAAA AATAACTATA ATCATTTTTTC AAAGTTAAAA 3960
 ATTTTGAATT TATGGTTAAC ATGTCAACTT ACTATGTGTA TAATGGTAAA CATTGATATT 4020
 15 AACTATATGT ATAAAAATGT CACGCAGATG CTATTTAAAT GTGATAAATA TTTTtagagg 4080
 TGAATAGAGT GGCTATAAAG CTAAGTTCAA TTGACCAATT TGAACAGGTT ATTGAGGAAA 4140
 20 ATAAATATGT TTTTGTATTA AACATAGTG AAAGTTGTCC AATATCGGCA AATGCGTACG 4200
 ATCAATTTAA TAAATTTTTA TATGAACGCG ATATGGACGG TTATTATTTG ATTGTCCAAC 4260
 AAGAACGCGA TTTGTCAGAT TATATTGCTA AAAAAACGAA CGTTAAACAT GAATCACCTC 4320
 25 AAGCATTTTA TTTTGTAAAT GGTGAAATGG TTTGGAATCG AGACCACGGT GATATCAATG 4380
 TGTCGTCATT AGCACAAGCA GAAGAATAAT GAACTATAG GTTTGAACA TTTTGCCTTA 4440
 CACTACTAGA CGTGAATAGC ACAACTTAA TCGTGTGAA TCAGAGTAGT TTGGCTATAA 4500
 30 TGATGTTCTG ACCTTTTATT TTATGTCACC TTTAGAAGCA GTTAAGTTAG TACTTTTTTTA 4560
 CAAACATATG TATAATATAT TCGAGTATTT TTATTGAAAa tATTTTGGAA AACGACGAAT 4620
 CCAATAAGAA AATTTAAACA TGATTTGTAA GTTAGTTTAA TAGGAAATAT ATGCTAAACC 4680
 35 AAAAGAAGCA TATTGTTATT TACTGGAATA ATTAATAATC ATGTCATGTT AAATGTTAGC 4740
 ATATAATCAC GAGATAAAAT CTAAAATTTA AGATTAATCT TTTATGAATA AAAAACGTAT 4800
 CACAACAAAT AATAAAGTAA GGTGGTCAAG GTTATGAAAG TATTAGTAGC CATGGATGAG 4860
 40 TTTCATGGAA TTATTTCAAG TTATCAAGCT AATAGATATG TTGAAGAGGC AGTTGCAAGC 4920
 CAAATTGAAA CTGCAGATGT AGTTCAAGTA CCATTGTTTA ATGGAAGACA TGAATTATTA 4980
 45 GATTCTGTAT TTTTATGGcm ATCTGGGcaA AAGTATCGTA TACCAGTACA TGATGCAGAT 5040
 ATGAATGAAG TTGAAGGTGT TTACGGACAA ACTGATACAG GGATGACCGT TATCGAGGGG 5100
 AATTTATTTT TAAAAGGTAA AAAACCAATT GTTGAACGAA CAAGTTATGG TTTAGGAGAA 5160
 50 ATGATTAAAC ATGCATTAGA TAACGACGCA AAACATGTTG TAATTTCACT AGGTGGGATT 5220
 GATAGTTTTG ATGCTGGTGC AGGTATGTTA CAAGCATTAG GTGCTCAATT CTATGATGAC 5280

55

	GATATGTCGA ACTTACACCC TAAAATGGAA ACAGCAAGAA TTCAAGTAAT GTCGGATTTT	5400
	TCAAGTCGAT TATATGGTAA GCAAAGTGAA ATCATGCAAA CTTATGATGC GCATCAGTTG	5460
5	AATCATAATC AAGCAGCAGA AATCGATAAT TTAATTTGGT ATTTTAGTGA GTTATTTAAA	5520
	AGTGAATTGA AAATTGCAAT TGGTCCAGTT GAACGTGGTG GTGCTGGTGG TGGAAATTGCA	5580
	GCAGTCTTGA ATGGACTGTA TCAAGCTGAA ATATTAACCA GTCATGCATT AGTAGACCAA	5640
10	CTAACACATT TAGAAAATTT AGTTGAACAA GCGGATTTAA TTATTTTGG AGAAGGATTA	5700
	AATGAAAATG ATCAGTTGCT AGAAACGACA ACATTGCGTA TTGCAGAACT TTGTCATAAA	5760
	CATCAAAAGG TTGCCATTGC AATTTGTGCA ACTGCTGAAA AGTTTGATTT ATTTGAATCA	5820
15	CAAGGGGTTA CAGCAATGTT TAATACATTT ATCGATATGC CAGAACTTA TACTGACTTT	5880
	AAAATGGGtT ACAAATTAGG CATTATACGG TTCAGTCTTT AAAACTGTTG AAAACACATT	5940
20	TTAATGTTGA GGTTTAGTAA AGAAGGACTA AATTGGTGAT GCTGTCATGA TGGTTAATAA	6000
	CATTTATGAT GGTTAGCAAA ACGAATTAGA AGATCGAAAG TATACGTAAA AAATATGAAA	6060
	AATCAGCTA TCATTGCACT GAATGTTAGC GTGATTTTTA TATATTAATT AAGCCTGAGT	6120
25	TGAACTAGTA TATAATCGTT GGTTTTTAGT GATTTTCAGC GATATCTTCT ACAATTCCAA	6180
	TGATTACTTG TACTGCTTTT TCCaTAACAT CAATGGATGC aTATTCATAT GGGCCGTGGA	6240
	AGTTACCGCA ACCTGTAAAG ATGTTTGGAG TTGGTAACCC CATAAATGAC AATTGTGAAC	6300
30	CATCTGTACC ACCGCGAATA GGTTCAGTGT TTGCTGGAAT ATCTAATTTG GCAAAGACAC	6360
	GTTTAGGTAT ATCAATAATA TGAGGCAATG GTAATATTTT TTCTGCCATA TTGAAATATT	6420
	GATCCGATAT ATCAACTTTA ACTGGATAAT TTTCAAATG GGCATTGATA TCGTCACGTA	6480
35	TTTCTAAAT ACGTTTCTTA CGCAATTGCA ATTGTTTTTT ATCATGATCA CGAATAATGT	6540
	ATTGCAAAGT TGCTTTTTCA ACAGTTCCTT CAAAGTTCAT TAAGTGATAA AAGCCTTCGT	6600
40	ATCCTTCTGT TCGCTCCGGA ACTTCACTAT CAGGTAGCAA ACTATCGAAT TGTTACCTA	6660
	AACGTATTGC GTTTACCATT GCATTTTTAG CTGAACCAGG ATGAACATTT ACACCGTGGC	6720
	ATGTAATAAC CGCTTCAGCA GCGTTAAAGC TTTCATATTG TAATTCTCCA TATTGACTAC	6780
45	CATCCATAGT ATAAGCAAAA TCAGCATTGA AGCGGTCAAC ATCAAATTTA TGTGGACCAC	6840
	GACCGATTTT TCGTCTGGT GTAAATCCAA TGCGAATGGT ACCATGTTTA ATTTCTGGAT	6900
	GTTCTTGTA ATAACAAATA GCTTCCATAA TTTCCACAAT ACCCGCTTTA TCGTCTGCAC	6960
50	CTAGTAACGA TGTACCATCA GTTACCATTA ATGTATGACC AACTAACTG TTAAGTTCTG	7020
	GAAATACTTT AGGATCTAAG ACACGTTTAG TATTGCCTAG TTTGTATGGC TTACCATCAT	7080

55

	GCGCCAAAAA TCCAACTGTT GGGACGTCGA CATCGATGTT ACTTTCTAAT GTAGCAAATA	7200
	AGTAGCCATT TTCATCTAAA TCAGTTGGCA ATCCTAATTG TTGTAATTCT TTTTCTAATA	7260
5	AATGTAACAA ATCCCATTGC TTTTCAGTTG AAGGTGTTGT TGTAGATTTT GGATCAGATT	7320
	GCGTATCAAT TGTCGTATAT CTTGTTAATC TATCTATCAA TTGGTTCTTC ATTATATTCG	7380
	ACCCCTTAAA CTCTATTATT CATGTTGTAA GATTTTTTAT ATGTCTTACC TTTGATTTTA	7440
10	CCATACAGTT GTTTGATACG TGTGTATAGG TAATATAGAA TTTCAGAAAC TAATATACCG	7500
	AAAGCAATCG CACCTGAAAT CAGTGTA CTT CTAAAAATGT ATTTACAGCA CTTGTATAAT	7560
15	CATTGATAC TAAAAACGA GTCGCTTGAT AAGCTGCACC ACCAGGTACT AATGGTATAA	7620
	TGCCTGGCAC TATGAATATA ATTACCGGTC GTTTATATCT GCGACTCATA GTATGACTCA	7680
	TTAAGCCTAA AATTAAGCTT CCCAAAAATG AAGCGCCAAC TTTTCCAAAC TCTAAATCTA	7740
20	CCGTTAATTG GTAAATCGTC CATGCAATGG CACCCACAAA TCCACATGCT ACTAAGAGGC	7800
	GTTTGGGTGC ATTGAAAATG ATAGAGAAAA G TA CTGTTGA TATAAAGCTG ATTGTAAAAT	7860
	GAAATAAATA AAATAGCATG CTTTAACAGT CCTTCCTTAA ATGATTAATA AAACGATTGC	7920
25	GACACCAGCA CCGATTGCGA ATGCTGTTAA TGCAGCTTCA ACACCGCGAG ACATACCTGC	7980
	AAGTAATTCA CCCGCTAATA AATCTCGAAT GGCATTGGTA ATTAATATAC CAGGGACAAG	8040
	TGGCATGACA CTGGCTATAG TAATGATATC TTGATTGGTT GCAATGCCTA ATTTAGTAAA	8100
30	TGTGGCTGCA ATGGATATGA CCACAGCGGC TGCAACAAAC TCTGAGAAAA ATTTAATTTG	8160
	TATATAGCGT TGCACAAAGC TGAATGTTAA AAATGCGGAT CCGCCAGCAA TGA CT GCAT	8220
	CCAACAATCT GATGCGACAC CACCAAACAT AAATAGGAAG AAGCCACATG CAATGGCAGC	8280
35	TGCAAAAGAA TTCGTTAAAA AAGAATATTG TAATGATGCA TGCTGTAAAT GAATAAATTC	8340
	AGATTTAGCT TCATCAATTG TGAGTTCTTT ATTTGATATT TTACGTGAAA GACTATTCGT	8400
40	TAAAGCGATT TTCTCTAAAT CTGTTGTACG CTCTTG T ACA CGAATTAATC TTGTACTTGT	8460
	TCGATCGTTT AATGAAAAAA TAATTGCAGT TGA ACT GACA AA CT TATATG TATTATGAAG	8520
	ACCATAACTA TGTGCGATAC GGTTCATTGT ATCTTCAACT CGATATGTTT CAGCACCTGA	8580
45	TTCaAGTAAA ATTCTACCTG CAATTAATAC AACATCAATC ACTTTGTTTT CATCTATAAT	8640
	TGTGATTGAA TCTGGCATAT CAATTCACCT CCAATGATAT GTGTTATTTA TTTGAACAAT	8700
	TGaAGTTTAC AACTTGTTGT TACAAC TTT C AATAGTGAGA CTTTGTGTTA GTATGATGAA	8760
50	CTTGTATGGT TCAAA TTT TAA ATAAGAAAAA CTGTTAATCT TTGCTATTAT ACTATGATTT	8820
	AATAATAGCA AAGGATTAAC AGTTTTGTCG TTGTTATAAA TTGATAATAG GGTTAAACAT	8880

55

TTTACGCTGT GATTTTGGAT CGTCATCTGT TAAATAACCA ACACCGATAG AACTGACAA 9000
 TTTAATAACT TCTTTGTTTG GTAAATGGAA TGATGATTTT TCAACACCCG AACGAATATT 9060
 5 TTCAGCTAAT TTAACACTTT GATCAAGTGA ATAATTGTGA ATGACAACTG AGAACTCTTC 9120
 GCCACCATTT CTAAAAATTT TAAATTGATT CGGCACATAG TTTTAAAGTA ATTGAGACAT 9180
 TTGTTTTAAT ACAGCATCAC CTGATTTGTG TGAGTAGGTA TCATTGaCAT CTTTAAATCC 9240
 10 ATCGATATCG ATTAATAATA ATGCGATACT TTGATGTTCT TTTTCAGCTT TTCGTGAAAT 9300
 TTCATTTAAA TGTCTATCAA ATTCTTTTAC ATTACCTAAG CCTGTTAAGT AATCATATTT 9360
 ATCTTCGTTT TCATAACGAT TTACGAGTGA GAAGAAATGC CAAATATCGA CAAATGTTAT 9420
 15 CGCTGAAGCT AAAGTGATAA TTAATGAAAT TGGTATTAAA ATGATAACTT CCGATAGTGT 9480
 GTAAATAGGA CTCACTAACG CGACACCAAA TAAATGATT ATTGTAACAA CATTAAGTAT 9540
 TAATAATGAT AGCACATCAT TTTGTTTTAA AAATGGTCCA ATAGCACTTG TTAGTGCAGC 9600
 AATAACAATC AACGTAACAC CGTACATAAT CGAGTTGTTA AATACTACAA TTTCAACAAT 9660
 TGCTACAATT ACTGTGGCAG ATAATGTATA GACCATATTT GTAAATCTAC CTAAAAACAA 9720
 25 TAAAGGAACG AATGTTAAGT GAATTAAATA ATCTTCACGA TAAGGGATAG GGTAGACAGA 9780
 TAATAATAAT GATACGATTG TCATTAAAAC AGTGACATAA GCCTTAGAAA AAAC 9834

(2) INFORMATION FOR SEQ ID NO: 38:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23439 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

40 TCTCAATCAG ATGAAAAATT GCATATCGTA GGTTTACAG AAAGTGCAAA ATATAATGCG 60
 TCATCAGTCA TTTTCACGAA TGACGCTACC ATTGCCAAGA TCAATCCTAG ATTGACTGGA 120
 GATAAAATTA ATGCAGTTGT TGTACGTGAT ACAAATTGGA AAGACAAAAA ATTAAACCAA 180
 45 GAGCTTGAAG CGGTAAGTAT TAATGACTTT ATTGAAAAATT TACCAGGTTA TAAACCACAG 240
 AACTTAACAT TAACTTTAT GATTTCATTC TTATTTGTCA TTTTCAGCTAC AGTTATAGGC 300
 ATTTTCCTAT ATGTCATGAC ATTACAAAAG ACGAGTTTAT TTGGCATATT AAAAGCTCAA 360
 50 GGATTTACGA ATGGCTATTT GGCGAATGTG GTAATTTTCG AGACGGTCAT ATTAGCACTA 420
 TTTGGTACGG CATTTGGCTT ACTGTTAACA GGCGTTACAG GTGCATTTTT ACCTGATGCA 480

55

	TCTGTATTAG GAAGTTTATT CTCCATTTTA ACAATTAGAA AAATAGATCC GTTAAAGGCG	600
	ATTGGGTAGG AGGTGTAGCA AATGTTGAAA TTTGAAAATG TAACAAAGTC ATTTAAAGAT	660
5	GGGAATCGTA ACATTGAAGC GGTAAAGAT ACAAATTTTG AGATAAATAA AGGTGATATT	720
	ATAGCATTGG TTGGACCTTC TGGCTCTGGT AAAAGTACAT TTCTAACTAT GGCAGGTGCT	780
	TTACAAACAC CGACATCTGG GCACATTTTA ATCAATAACC AAGATATTAC GACAATGAAG	840
10	CAAAAAGCAT TGGCAAAAGT TAGAATGTCT GAAATAGGTT TTATTTTACA AGCTACAAAC	900
	CTTGTACCAT TTTTAACGGT AAAGCAACAA TTTACATTAT TGAAAAAGAA AAATAAGAAT	960
	GTTATGTCTA ATGAAGACTA TCAGCAACTT ATGTCACAAT TAGGTCTAAC TTCATTGCTT	1020
15	AATAAGTTAC CTTCAGAAAT TTCAGGTGGT CAGAAACAAC GTGTGGCGAT AgCaAAGCGT	1080
	TATATACGAA TCCGTCGATT ATTTTAGCGG ATGAACCTAC CGCGGCGTTA GATACTGAAA	1140
20	ATGCGATTGA AGTCATTAAA ATTCTACGTG ATCAAGCCAA ACAAAGAAAG AAAGCATGTA	1200
	TTATTGTTAC ACATGATGAA CGACTTAAAG CATATTGTGA TCGTTCATAT CATATGAAAG	1260
	ATGGCGTCTT TAATCTTGAA AATGAAACAG TAGAATAGTT TTATTAAGCC GGTACATCAT	1320
25	GTGCCGGTAT TTTTATGTTT ATGTATTATT TGAATAAACT TTCACATTCA ATTAATAATA	1380
	ATTATTATCG AAAATCAGAA ATATTCCGTG AAATATAATA TTTTTTGTAG TAAAATGGCC	1440
	TCTAAGTATT CAATATTTAA ATATGGGGAT TGAATATAAA ATTATCGTAA TGGGGGTCAA	1500
30	TGGTTATGGA TTTATTGATA GGTACTTTAT TTTTATTTT GGTCTTAGTG ATTTTACAT	1560
	TATTTACATA TAAAGCGCCT AATGGTATGC GTGCCATGGG AGCATTAGCT AATGCAGCAA	1620
	TCGCAACATT TTTAGTGGA GCATTTAATA AATATGTTGG TGGCGAAGTA TTCGGTATTA	1680
35	AATTTTTAGA AGAGCTAGGA GACGCTGCGG GAGGTCTAGG TGGTGTGCT GCGCTGGAT	1740
	TAACAGCATT AGCTATCGGT GTGTCACCAG TATATGCATT AGTTATAGCA GCCGCTGCG	1800
40	GTGGTATGGA TTTATTACCA GGTTTCTTTG CGGGTTATAT GATTGGATAT GTGATGAAAT	1860
	ATACAGAGAA ATATGTGCCG GATGGTGTG ACTTAATTGG ATCGATTGTC ATCTTAGCGC	1920
	CATTAGCTCG TCTTATTGCA GTATTATTAA CGCCAGTAGT GAATAGTACA TTGATTGAA	1980
45	TTGGTGATAT TATCCAAAGT AGTACGAATA CGAATCCAAT TATCATGGGT ATCATTTTAG	2040
	GTGGTATTAT TACGGTTGTC GGCACAGCGC CATTGAGTTC AATGGCATTG ACAGCATTAT	2100
	TAGGTTTAAAC GGGTGTACCT ATGGCTATTG GTGCCATGGC AGCATTTAGT TCGGCATTTA	2160
50	TGAATGGGAC GCTATTCCAT CGCTTAAAT TAGGTGATCG TAAGTCTACG ATTGCAGTAA	2220
	GTATTGAACC TTTATCACAA GCAGATATTG TATCAGCCAA TCCAATTCCA ATCTATATTA	2280

55

	ATGCGACAGG TACAGCTACA CCGATTGCAG GATTTTTAGT TATGTTTGA TTTAATCATC	2400
	CGACGACAAT TGTGATTAT GGTGTAGTAA TGGCGATTGT AGGTGCGCTT GCAGGTTATC	2460
5	TTGGTTCAAT TGTATTTAAA AAATATCCAA TTGTTACTAA GCAAGACATG ATTAATCGAG	2520
	GTGCAGTAGA CGCATAGCAT CATCATATTG AATAGTAAAA ACAATAAAAA CATAGTAACG	2580
	TGATTCACTC GATGTAACAG TCGATAATGA GTCACGTTTT TTTATAGAAA AATACAAGAC	2640
10	ATAAAAATGT CATAATTTAT TGTCGACAAA TATCATACTG TATAAACATT TATCATTTTC	2700
	TCAAGTACCT TTTACACGAT GGAATGAACT TACTTTTTAC GAAATTATGC GTATTTTATA	2760
	AACAAATATC ATTGATATAA CGGTAAATGT AAGCGTTTAC AACAGAAATA ACAGCATGCT	2820
15	ACGATATTTT TGTAAATTCA CTGATTCAAG TATTTTAAAGT CAATATGAGG AGGGATGTTA	2880
	TGAGCGATTG TGAGAAAAGAA ATTTTAAAAA GAATTAAAGA TAATCCGTTT ATTTCAACAAC	2940
20	GTGAACTTGC TGAGGCAATT GGATTATCTA GACCCAGCGT AGCAAACATT ATTTCAAGAT	3000
	TAATACAAAA GGAATATGTT ATGGGAAAGG CATATGTTTT AAATGAAGAT TATCCTATTG	3060
	TTTGTATTGG CGCAGCGAAT GTAGATCGTA AGTTTTATGT GCATAAAAAT TTAGTTGCAG	3120
25	AAACATCAAA TCCTGTAAACG TCAACACGCT CTATTGGTGG CGTAgCAAGA AATATTGCTG	3180
	AGAACTTAGG TAGGCTTGGC GAAACGGTCG CTTTTTATC TGCTAGTGGA CAAGATAGTG	3240
	AATGGGAAAT GATTAAACGA TTGTCCACAC CATTTATGAA TTTGGATCAT GTTCAACAAT	3300
30	TTGAAATGC GAGTACAGGT TCATATACAG CTTTAATTAG TAAAGAAGGC GACATGACAT	3360
	ATGGCTTaGC AGATATGGAA GTGTTTGA CTACATTACGCC TGAATTTTTA ATTAAGCGTT	3420
	CACACTTATT GAAAAAGGCT AAGTGCATTA TTGTAGATTT GAATTTAGGC AAAGAGGCAT	3480
35	TAACTTCTT ATGTGCCTAT ACCACGAAAC ATCAAATCAA ATTAGTTATC ACCACGGTTT	3540
	CTTCCCCAAA AATGAAAAAT ATGCCTGATT CATTACATGC TATTGATTGG ATTATCACGA	3600
	ATAAAGATGA AACAGAAACA TACTTAAATT TAAAAATAGA ATCTACTGAT GATTTAAAAA	3660
40	TAGCTGCTAA ACGCTGGAAT GATTTAGGTG TTAAAAATGT TATTGTGACA AATGGCGTGA	3720
	AAGAACTCAT TTATCGAAGT GGTGAGGAAG AAATCATTAA GTCAGTTATG CCATCAAATA	3780
	GTGTGAAAGA TGTTACAGGT GCAGGCGATT CATTCTGTGC TGCAGTAGTG TATAGCTGGT	3840
45	TAAATGGGAT GTCTACTGAA GATATATTAA TTGCTGGTAT GGTTAACGCA AAGAAAACGA	3900
	TAGAAACGAA ATATACAGTT AGGCAAAACC TAGATCAACA GCAACTTTAT CACGATATGG	3960
50	AGGATTATAA AAATGGCAAA TTTACAAAAG TATATTGAGT ATTCTCGAGA AGTTCAGCAA	4020
	GCACGGGAGA ACAATCAACC GATTGTAGCA TTAGAATCAA CAATTATTTT GCATGGTATG	4080

55

	GCCATTCCAG CAACCATAGC CATTATAGAT GGCAAAATTA AAATTGGTTT AGAAAGCGAA	4200
	GATTTAGAAA TACTGGCAAC TAGTAAAGAC GTTGCTAAAG TATCTAGAAG GGATTTAGCA	4260
5	GAAGTTATTG CGATGAAGTG TGTGGTGCT ACTACTGTAG CGACGACGAT GATATGTGCT	4320
	GCAATGGCTG GTATTCAATT TTTTGTTACA GGAGGTATTG GGGGCGTCCA TAAAGGTGCA	4380
	GAACATACGA TGGACATTTC AGCAGACTTA GAAGAACTGT CTAAAACAAA TGTCACTGTT	4440
10	ATCTGTGCAG GTGCCAAATC AATTTTAGAC TTACCTAAGA CGATGGAGTA TTTAGAAACA	4500
	AAAGGCGTTC CAGTTATTGG ATATCAAACG AATGAATTGC CAGCATTCTT CACTCGCGAA	4560
	AGCGGTGTTA AGTTAAACAAG TTCGGTTGAA ACGCCAGAAC GACTTGCTGA CATTCAATTA	4620
15	ACAAAACAGC AGTTAAATCT TGAAGGTGGC ATTGTTGTTG CTAATCCAAT TCCATATGAG	4680
	CATGCCTTAT CAAAAGCATA TATTGAGGCA ATCATAAATG AAGCTGTTGT TGAAGCGGAA	4740
20	AATCAAGGTA TTAAAGGTAA GGACGCCACA CCGTTCTTGT TAGGGAAAAT TGTAGAAAAA	4800
	ACGAATGGTA AAAGTTTAGC AGCAAATATA AAACCTGTTG AAAACAATGC GCGTTGGGT	4860
	GCTAAAATTG CTGTCGCTGT TAATAAATTA TTGTAGGTGA TGATACATGA ATATTTTATT	4920
25	CGCTATCACA GGGATAGCAT TTGCACTATT TGTGCGTTT TTATTCAGTT TTGATCGTAA	4980
	AAAAATAGAC TTCAAAAAGA CGTTAATAAT GATATTTATT CAAGTGTGA TCGTGTATT	5040
	TATGATGAAC ACAACGATTG GTTTGACAAT TTAACTGCA CTAGGTTTAT TTTTGAAGG	5100
30	GCTAATAAAT ATTAGTAAAG CAGGCATAAA TTTTGTTTT GGAGATATAC AAAATAAAAA	5160
	TGGCTTTACG TTCTTTTAA ACGTATTACT GCCATTAGTT TTTATTTCTG TATTAATAGG	5220
	CATCTTTAAT TATATTAAGG TATTACCATT TATTATCAAA TATGTAGGTA TCGCTATTAA	5280
35	TAAATAACT AGAATGGGGC GCTTAGAAAG TTATTTTGCT ATTTCAACAG CAATGTTTGG	5340
	GCAACCGAGAA GTATATTTAA CAATAAAGA TATTATTCCA AGATTATCTA GAGCGAAATT	5400
40	ATATACAATT GCGACGTCTG GTATGAGTGC TGTTAGTATG GCAATGCTAG GTTCATATAT	5460
	GCAGATGATT GAACCCAAGT TCGTAGTTAC AGCAGTAATG TTAAATATTT TTAGTGCCT	5520
	TATCATCGCC AGTGAATCA ATCCCTATAA ATCTGATGAT ACTGATGTTG AAATTGATAA	5580
45	CTTAACGAAA TCCACAGAAA CTAAAACATT GAATGGAAAA ACAGGAAAAC CTAAGAAAGT	5640
	TGCCTTTTTC CAAATGATTG GTGATAGTGC GATGGATGGG TTTAAAATCG CTGTTGTAGT	5700
	AGCCGTAATG TTGTTAGCAT TTATTTTATT AATGGAAGCA ATTAATATCA TGTTTGGTAG	5760
50	TGTTGGTTTG AACTTTTAAAC AGCTTATTGG CTATGTGTTT GCACCAATCG CATTCTTAAT	5820
	GGGGATTCCA TGGAGCGAAC TGTTCCAGCT GGCTCTTTAA TGGCGACTAA ATTAATTACA	5880

55

	CAAGGTATCA TTTCAGTTA CTTAGTAAGc TTCGCTAATT TTGGTACGGT TGGTATCATC	6000
	GTAGGTTCAA TTAAAGGCAT TAGTGATAAA CAAGGAGAAA AAGTTGCATC CTTTGCAATG	6060
5	AGGTTGCTAC TTGGTTCAAC TCTAGCTTCA ATCATTTTCAG GATCAATCAT TGGCTTAGTA	6120
	TTGTAAATGA ATCGAAGTAC CTAAATTAAA TTCATGGCAA AGCTAAACCC CGTCACCAAG	6180
10	TTGGCGCAAC AGCGcATgcA TAACTTAGTG ACGGGGTTTT ATCATAACAA TCTACTTTTT	6240
	CGTAGCCGTT TTTGAAATGT ATGTTGATGG TTTATCTTTT TCAAAAATTG TTAATCCCGT	6300
	TATATCTTTT TTATGTTTTG AAGGGACAAT GAAGCTAAGT ATATAAGCAA AGACAAAAGC	6360
15	AACTGTAAAT GAAATGGTAG ATACATAGAA AGGTGAGTTA CCTTTGCCAA CACCATTATA	6420
	GACATAAGCA AAGATGATAC CCAATATTAA TCCACAAATA ACACCGAATG TATTCGTACG	6480
	TTTAGTGAAA ATACCAACTG CAAATACACC AGCCAATGGA ACGCCGAATA ATCCAGTCAC	6540
20	AAACAAGAAAT AAATCCCATA AGTCATTTGA ATTAGAAGCA ATTAAGTATA GTGACATTCC	6600
	AAAACCGAAA ATACCTGCAA TGATAATAAT GAAACGTGCA AAGTTAACTT CGTGTCGCTC	6660
	GCTACCTTTT CCGAAGAAGC GTTGCTTAAT GTCGATTGAA ATACAAGCAG ATATAGAATT	6720
25	TAACTAGAT GAAATGGTAG ACTGTGCAGC GGCGAAAATG GCTGCAATAA GTAATCCTGC	6780
	TACAAATGGT GGCATCTCAG TCAAAATGAA ATATGGCACT ACAGATGATG TATTGAAGCC	6840
	TTTTGGTAAA ACAGCTTCAT GTGTATAAAA TGAATACAGC ATTGTACCCA TACCATAAAA	6900
30	TAAGGGTGCT GAAATTAAAG CTAGGATACC ATTTGTCCAT AACGATTIAT TTGTTTCTTT	6960
	TAACTATCA GAAGCTTGAT AACGCTGCAC GACGTCTTGA CTCGCTGTGT ATTGATACAA	7020
	GTTGTTGAAA ATATTTCCCTA GGAAAATAAT TGGAATGGCA GCTGCCGAG TATTTAGTTT	7080
35	CCAATGTCT GCACTAATTA ATTTTTGTG CTCAATCGCA TCTGCAAAGA CAGTGCCGAA	7140
	ACCGCCTTTA ATGTTCAAA CACCTAGAAT AATAATAACT AAAGCGCCGC CTAATAAAAT	7200
40	GACGCCTTGA ATGAAATCAC TCCAAACCAC ACCTTCGAAA CCACCTAAAA ATGTATATAA	7260
	AATACATAGT AAACCAACGA GTGATGCAAC GATATAAGGG TTCATGTCTG ATACAGATGT	7320
	GATTGCTAAT GTTGGTAAGT AGATAACAAT TGCAACACGC CCTAAATGGT AAACGACAAA	7380
45	TAATAATGAG CCAATGACAC GTATGCTAGG GCCAAATCTA GCTTCTAAAT ATTCATATGC	7440
	AGATGTTACC TTTAACTTTT TAAAGAAAGG GACATAGAAA TAAATAAGTA ATGGAATAAT	7500
	TGCGACGATA GCAATGTTAC CAGCGATATA TGACCAATCT GTTAAAAATG CTTTCTCTGG	7560
50	TGTCGACATA AATGTAATCG CACTTAACGT AGTAGCATAA ATTGAAAAGC CAACTACCCA	7620
	AGATGGCAAG CGACCACTTG CGGTAAAGAA ACTATTGGTA CTTTGGCTCG CGCGCTTGGT	7680

55

	TGTGCCAAAT CCAACTTCTT TCATGGGCAA CATCCCCTTT ACAATGTATT GATTCTTTGA	7800
	TGTCTATAAA TCGTATTTTG CAATGAGTTG ATCTAATGTT TGTCGATGTG CTTCGTTAAA	7860
5	AGGTTTGAAA GGTCTTTTCG GTAATCCTGC ATCAATGCCA CGATGACGTA ATATTTCTTT	7920
	CAATGTTGGA TAAATCCCCA TTGATAACAC TGTTCGATA ATGTCGTTTG AATCATGTTG	7980
	CAGTTGGTAA GCTTCTTGAA TTTGACCTTG TCGTGCTAAG TCGAAGATTT TTCTTGACG	8040
10	GCGACCAITTA ACGTTATATG TAGAACCAAT TGCACCATCT ACGCCAGAAA TCGTAGCTTG	8100
	AACTAACATT TCATCAAAGC CAGATAAGAT TAATTTGTCT GGAATGCTT TTCTAATACG	8160
	TTGAGTAGG AAGAAGTTTG GCGCTGTATA TTTAACACCA ACAATTTTTT CATGATTAAA	8220
15	TAGCTCGCTG AATTGTTCAA TAGAAATATT CACACCTGTT AAATCTGGTA TTGCATAAAT	8280
	AATCATATTG TTCTGAGTTG CTTCGATAAT ATCGAAATAG TAATCTCTAA TTTCTTCAAA	8340
	AGTAAATGGA TAGTAGAATG GTGTTACGGC AGAAAGTGCA TCATAACCGA GTTCTGTGGC	8400
20	ATATTTTCCA AGTTCAATGG CTTCATTTAA ATCTAACGAA CCTACTTGAG CAATCAATTT	8460
	CACTTTATCC CCAACTGCCT CTTTGGCAAC CTTGAAAACT TGCTTCTTCT GCTCTGTATT	8520
25	TAATAAAAAG TTTTCGCCGT AGCTACCATT TACATAAAGA CCGTCTAATT CTTCAGTTTC	8580
	AATGGCATTT TGAGCAATTT GTTTAAGTCC TTGTTCAATTT ACTTGACCAT TTTCATCAAA	8640
	AGGAACGAGT AACGCTGCAT ATAAACCTTT TAAATCTTTG TTCATTATGA AGTCCCTCCA	8700
30	AAAATCATTT GATAATATAG TTTACAGCTA TAATTGTAAG CGCTATCATA AAATGTAACA	8760
	ATATCTTTTT GAAAATTGTA GTCATATTTA TGTATAATTA ATGAAAATGT TTTTCAAAAT	8820
	CAATAGAAAT GGAGTGAGTA AGGTGTATTA CATCGCAATC GATATTGGAG GCACTCAAAT	8880
35	TAAATCGGCA GTTATTGATA AGCAATTGAA TATGTTTGAC TATCAACAAA TATCAACGCC	8940
	GGACAACAAA AGTGAGCTTA TTAAGTACAA AGTATATGAG ATTGTAACAG GATATATGAA	9000
	GCAATATCAG TTGATCCAAC CTGTCATAGG TATTTTCATCA GCAGGCGTTG TTGATGAACA	9060
40	AAAAGGCGAA ATTGTATACG CAGGGCCAAC CATTCCGAAT TATAAAGGTA CTAATTTTAA	9120
	GCGATTATTA AAATCACTGT CTCCTTATGT CAAAGTAAAA AATGATGTAA ACGCTGCATT	9180
	ACTAGGCGAA TTGAAATTAC ATCAATATCA AGCAGAACGG ATCTTTTGTA TGACGCTTGG	9240
45	TACAGGCATT GGGGGTGCGT ACAAGAATAA TCAAGGTCAT ATTGATAATG GTGAGCTTCA	9300
	TAAGGCAAAT GAAGTTGGGT ATTTATTGTA TCGTCCAAC TGAATAACAA CGTTTGAGCA	9360
50	ACGTGCTGCA ACGAGTGCAT TGAAAAAGCG CATGATTGCC GGAGGATTTA CGAGAAGCAC	9420
	ACATGTGCCA GTATTGTTTG AAGCAGCTGA AGAAGGTGAT GATATTGCAA AACAAATATT	9480

55

	AGGGCTTATA TTAATTGGGG GCGGTATATC TGAACAAGGA GATAATCTCA TTAAATATAT	9600
5	CGAGCCGAAA GTTGCACT ATTTACCAA AGACTATGTT TATGCACCAA TACAAACGAC	9660
	TAAGAGTAAA AATGATGCAG CATTATATGG CTGTTTGCAA TGATAGTTGA AAGAAGGAGT	9720
	CATTCTAAAA TAGAATTGTA AACCGTTACG AGAGATGAGA GCTGTTGTTA GTTCCACACA	9780
10	TCACACTCTA TCTAGGACCA ATCTAAACTA TATCAACCAA CAGTGTGCCA CGGGCAAATT	9840
	AAATTGAAGA AGCTGAGATA TTAAAATTTT AGAAAATGTA AAAAAATATT TGGTATTGAA	9900
	ATTAAAAAAG CACCTAGCAA CTCGTTGGGA CAATCACGAT GATTGTCTAC AGTTGCAGGT	9960
15	GGATTTGAAT ATACTACTAG TTATTTGTTG TCTAGGATAA TAGATTTAGT ATGTTGATAA	10020
	GTTTGACTCA GATTCGTATT TTCTAATAAA TGATAACTCA CGATATCGAT TAAAAAGAGT	10080
	GTCGCAATTT GTGTGTTGAT AAATTGATGG TCGGTATTAC GCGATTGATC CGTTGTTAAA	10140
20	AGTACTAAAT CTGCACAATC TGTAAGTTTA CTACCTTCAA AATTTGTGAT GGCAACGACA	10200
	TATGCACCAT GAGATTTGGC GACTTCCGCT GCAGAAATTA ATTCCGAAGT ATTACCACTA	10260
	TTTGACATAG CAATAAACAT ATCCGAATGA GATAGTAGGG ATGCCGATAT TTTCATTAAA	10320
25	TGTGAATCGG TAGTAACATT ACCTTTTAGC CCCATACGAA TCATACGATA ATAAAAATCA	10380
	GTCGCTGATA AACCAGAGCT ACCTAGTCCA GCAAAGAGTA TATGTCGACT TGATTGAAGT	10440
	TTGTCGATAA AGGTTTGAT AATGTCGTTA TCAATAAATT CACCAGTTTG TTGAATGATT	10500
30	TGTTGATGAT ATTTATGAAT TCTTTGAATA ATTGGGCTAT TTTCAATAAC TGTCTCTGTC	10560
	ATTTCTTGTT GAATATTAAT TTTTAAATCT TGGAAATCT CATAATCCAG CTTATGACTA	10620
	AAGCGTGTC TCGTTGCTGG TGATGTACCA ATCGCATGGG CTAAGGAGTT AATCGTTGAA	10680
35	AAGGCATCGC TATAACCATT TTGTCTTATA TAATTGACGA TCGGTTTATC AGTTTTTGTA	10740
	AATTAATGTT GATAACGTTG AACACGATTC TCAAAATTCA TTGTGTCACC CCTTCATCTT	10800
40	AATGATTACT ATTATATATG AAAAAATTTT TCAAGATAGT AAAAAGCATT GATAAAAATT	10860
	ATCTTAATGA TATATTGTAA ATGACTTTAC GTGAAAAAAC GACTTATGGA GTGAGGAATA	10920
	ATGTTACCAC ATGGATTAAT AGTATCTTGT CAGGCACTAC CAGATGAACC ATTGCATTCA	10980
45	TCTTTTATTA TGTCGAAAAT GGCATTAGCT GCGTATGAAG GTGGTGCTGT TGGTATTCGC	11040
	GCAATACTA AGGAAGACAT TTTAGCAATT AAAGAAACGG TAGATTTACC AGTTATTGGC	11100
	ATTGTGAAAC GTGACTATGA TCACTCAGAT GTTTTCATTA CTGCAACGTC AAAAGAAGTT	11160
50	GATGAACTGA TAGAAAGCCA ATGTGAAGTC ATTGCATTGG ATGCAACGTT ACAGCAACGT	11220
	CCGAAAGAAA CGTTAGACGA ATTAGTATCA TATATTAGAA CACATGCACC GAACGTTGAA	11280

55

	TATATTGGCA CGACGTTACA TGGCTATACT AGTTATACGC AAGGACAATT ACTTTATCAA	11400
	AATGACTTCC AATTTTTTAAA AGATGTACTA CAAAGTGTTG ATGCAAAAGT TATTGCGGAA	11460
5	GGTAATGTCA TTACACCGGA TATGTATAAA CGTGTGATGG ACTTAGGCGT TCATTGTTCA	11520
	GTCGTTGGTG GTGCGATAAC ACGACCAAAA GAAATTACGA AACGTTTTGT TCAAATTATG	11580
	GAAGATTAAA TGATAACGAT AAAAAACGA GATGACCATC ATTAATTAAA GGCACCTAAT	11640
10	TATCTTAGGT GGCTGAATGA ATGTAATGGG TTCATCTCGT TTTGTTTGTT TATGATAGTG	11700
	ATTTTATTTT CAACTTTATC CAAAAATAAG TAAAGCGACG GGGATGGTGA TTAATAGCGA	11760
	CAACGCCACG CGTAAAAACC AAATGATGAT GAGTTTCCAG ACAGGTATTT TAATTTCACT	11820
15	TGCTAGTATA CATGGCACTA ATGCTGAGAA AAAGATAATG GCTGATACGC TTAATACACC	11880
	GACGACAAAT TTAGTACTCA TTGCAGCTTT AGTTACTAAC AAAGATGGTA GAAACATCTC	11940
20	TACAATAGAA AaCTGACGC TTTTGCTAGT AAAGCCTGAT CAGCAATTGG GAAATATAA	12000
	ATAAATGGAT AGAAGATATA GCCAAGCCAA TCAATGAATG GTGTATAGTT CGCTACAATC	12060
	AGTCCTAAAA AACCAATCGA TAATATAGAA GGTAAATAC CAACAGTCAT TTCTAAACCG	12120
25	TCTTTCAAAT TGTCCCAAAC GTTCTTCACG AGAGATGGTG TTAATGCATT TTGTTTCATC	12180
	GCCTCTGCAT ATGCAGTTTT CAGTCTGCTT CTTTCAATAG CAACTTCTTG TTCTCCTTCT	12240
	TGTCCGTTAT AATATTCTGT TGATTCATTG CTGATTGGCG GTAGCCATGC AGTAATTGCA	12300
30	GTCACGACAA ATGTGATGAC TAAAGTTATC CAAAAGTATA AATTCCAATG CGGCATTAAT	12360
	CCTAAAGTTT TAGCAACGAT AATCATAAAA GTTGCTGAAA CTGTTGAAAA GCCAGTCGCA	12420
	ATAATCGTGG CTTCTCGTTT GTTGATACATC CTTGCTTAT AGACACGATT AGTAATCAAT	12480
35	AATCCTAAGG AATAACTGCC GACAAACGAA GCCACTGCAT CGACAGCGGA TTTTCCTGGT	12540
	GTTTTAAAA TAGGTCTCAT AATAGGCTCC ATATAAACAC CGACAAATTC TAATAAGCCA	12600
	TAGCCCACTA ATAAAGAAAG cGcAATTGCA CCTACTGGAA TTAAGATACT TAATGGCATC	12660
40	ATTAATTTTT CAAACAAAA CGGACCATAG TTAGCTTTAA ATAGTATTGA TGGACCGATT	12720
	TTAAATACAT ACATTATACC GATCATTGCA CCTGCAACTT TAAATAATGT AATGACCAAG	12780
45	TTTGTGATTG AAGTCATAAA AGTACGTCTC ACTATTGGTA ACGCTGTACC AATTAAAATC	12840
	ATAATCAGTG CAACATAGGG CATAAGTGGA CCTATGATTG AGCGAATGGC TAGATGAACA	12900
	TGATCGACGA AAATAGTGTT GTTACCATTA ATCGTAAAAG GAATAAAGAA ACATAGTATG	12960
50	CCCACTAAAC TATAGACAAA AAAACGCCAT GCACTTGTTT GTTGTGCATT AGAATGATAT	13020
	TGATTCATTA AAGCAACCCC TTTGTTTAAA TGAATACACA AAAGTGTATG ATGCATCTTC	13080

55

	ATAGTTTGAA TTATTTTCAT ACCAATACAA ATTAATAAT TATATATAGA TTGAACTAT	13200
	ATTACTTAAT AAAATATTTA TCTTAAATGT TGTGTGTTG ATTCAACACC ACAACTAAAA	13260
5	GTGTTTATAA ATTATTGGA AATACACATA TTTGTAAATG ATTAGTATCG ATTTAATATC	13320
	GTATTATTAA ATTTTATTAA ATTTTGTAGT CTTAATCmAA AAATAATATA TGTCATGTTA	13380
	TATTGAAGGT GCAGTTGTTT TTCATTCTCA AGAGGGGGTC AAAAAAATAC TTTTGAGGTG	13440
10	ATTATATGTT AAGAGGACAA GAAGAAAGAA AGTATAGTAT TAGAAAGTAT TCAATAGGCG	13500
	TGGTGTGAGT GTTAGCGCT ACAATGTTTG TTGTGTCATC ACATGAAGCA CAAGCCTCGG	13560
	AAAAACATC AACTAATGCA GCGGCACAAA AAGAAACACT AAATCAACCG GGAGAACAAG	13620
15	GGAATGCGAT AACGTCACAT CAAATGCAGT CAGGAAAGCA ATTAGACGAT ATGCATAAAG	13680
	AGAATGGTAA AAGTGAACA GTGACAGAAG GTAAAGATAC GCTTCAATCA TCGAAGCATC	13740
	AATCAACACA AAATAGTAAA ACAATCAGAA CGCAAAATGA TAATCAAGTA AAGCAAGATT	13800
20	CTGAACGACA AGGTTCTAAA CAGTCACACC AAAATAATGC GACTAATAAT ACTGAACGTC	13860
	AAAATGATCA GGTTCAAAAT ACCCATCATG CTGAACGTAA TGGATCACAA TCGACAACGT	13920
25	CACAATCGAA TGATGTTGAT AAATCACAAC CATCCATTCC GGCACAAAAG GTAATACCCA	13980
	ATCATGATAA AGCAGCACCA ACTTCAACTA CACCCCGTC TAATGATAAA ACTGCACCTA	14040
	AATCAACAAA AGCACAAGAT GCAACCACGG ACAAACATCC AAATCAACAA GATACACATC	14100
30	AACCTGCGCA TCAAAATCATA GATGCAAAGC AAGATGATAC TGTTCGCCAA AGTGAACAGA	14160
	AACCACAAGT TGGCGATTTA AGTAAACATA TCGATGGTCA AAATCCCCA GAGAAACCGA	14220
	CAGATAAAAA TACTGATaAT AAACAACATA TCAAAGATGC GCTTCAAGCG CCTAAAACAC	14280
35	GTTCGACTAC AAATGCAGCA GCAGATGCTA AAAAGGTTG ACCACTTAAA GCGAATCAAG	14340
	TACAACCACT TAACAAATAT CCAGTTGTTT TTGTACATGG ATTTTITAGGA TTAGTAGGCG	14400
	ATAATGCACC TGCTTTATAT CCAAATTATT GGGGTGGAAA TAAATTTAAA GTTATCGAAG	14460
40	AATTGAGAAA GCAAGGCTAT AATGTACATC AAGCAAGTGT AAGTGCATTT GGTAAGTAACT	14520
	ATGATCGCGC TGTAGAACTT TATTATTACA TTAAAGGTGG TCGCGTAGAT TATGGCGCAG	14580
	CACATGCAGC TAAATACGGA CATGAGCGCT ATGGTAAGAC TTATAAAGGA ATCATGCCTA	14640
45	ATTGGGAACC TGGTAAAAAG GTACATCTTG TAGGGCATAG TATGGGTGGT CAAACAATTC	14700
	GTTTAATGGA AGAGTTTTTA AGAAATGGTA ACAAAGAAGA AATTGCCTAT CATAAAGCGC	14760
50	ATGGTGGAGA AATATACCA TTATTCATG GTGGTCATAA CAATATGGTT GCATCAATCA	14820
	CAACATTAGC AACACCACAT AATGGTTCAC AAGCAGCTGA TAAGTTTGGA AATACAGAAG	14880

55

ATTTAGGATT AACGCAATGG GGCTTTAAAC AATTACCAAA TGAGAGTTAC ATTGACTATA 15000
 TAAAACGCGT TAGTAAAAGC AAAATTGGA CATCAGACGA CAATGCTGCC TATGATTTAA 15060
 5 CGTTAGATGG CTCTGCAAAA TTGAACAACA TGACAAGTAT GAATCCTAAT ATTACGTATA 15120
 CGACTTATAC AGGTGTATCA TCTCATACTG GTCCATTAGG TTATGAAAAT CCTGATTTAG 15180
 GTACATTTTT CTTAATGGCT ACAACGAGTA GAATTATTGG TCATGATGCA AGAGAAGAAT 15240
 10 GGCGTAAAAA TGATGGTGTC GTACCAGTGA TTTCGTCATT ACATCCGTCC AATCAACCAT 15300
 TTGTTAATGT TACGAATGAT GAACCTGCCA CACGCAGAGG TATCTGGCAA GTTAAACCAA 15360
 TCATACAAGG ATGGGATCAT GTCGATTTTA TCGGTGTGGA CTTCTTGAT TTCAAACGTA 15420
 15 AAGGTGCAGA ACTTGCCAAC TTCTATACAG GTATTATAAA TGACTTGTTG CGTGTTGAAG 15480
 CGACTGAAAG TAAAGGAACA CAATTGAAAG CAAGTTAAAT TCATCTTCTG AATTTAATAT 15540
 GCTATGTAAA TCGTGCTGTT ATCATGGCAC ATCAGATATA AGTAGCATCA CAGTGTTGAA 15600
 TTTAAAAATA GTAAAGTGAA ATAAAGCGCC TGTCTCATT GCGAAACTA AAGGGACAGG 15660
 CGTATCTGTT TATGAGCTTA ATAAATTGTA TGAATAATAT GGTTGATCGA ATAAGTGT 15720
 25 ATCATGATGA TAAATTGAGT TTTTAAAT AATGATATAT TACATCATTG TTATAGCGTT 15780
 TAAGAAATCA ACAACTTAC GATAAATAGT GATTGCTTCG TCATTAGGTC TACGATCAAA 15840
 ATCATGCTCG TTTTATTCA CGCGTTCAA TGTGAATGT GGAACATGAT TCATGATATG 15900
 30 TTCGCTTTCC TCAACGGGAA CATCATAATC GCCATTACAA TCGCAATGA AAACAGGTGG 15960
 AAGTGTTTTA AGTTCATCTG GTGCAATATT ATATTTTGAA TTAGTATAAT CAGCAATGTT 16020
 AATCATATTT ATCCATTTAC CTGTGCCACG TGCATAAACG TAGATTAAAA AACGTTGTGC 16080
 35 GATTTGATCT TGAACAACCG GTGTTGGTGA AGTGAGTTGT GCAATCATTG TTTCGTTTAC 16140
 GCTTTGAGCT ATTTTGCCT AATAACTATT AGTTGTTTTA AAAGGTTGAG TGTGATGCG 16200
 ACTATAACCA TAAAAATCAA TAACACCATC AATATCTCTG TCTCGTGCAA TTAATAGACT 16260
 40 TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT ATTGTGATTG 16320
 AATCGCATCG AATGAtGCgn AGnACATCCT CAATAATGCA ATCGAGACTT ACTTCTGGTA 16380
 ATAAACGATA ACTTAGTTGA ATTAAATCGT AATGTTCCGT AAgATATCGA TATACTGTGG 16440
 GGATAAATCG TTAGCTTTAC CGAACATTAA TCCACCACCG TGGATGTAGA CAATAGCGCC 16500
 TTTTGTGGT TGATTTTTTG CTTTAATAAT TGTGTAAGGT AATGCAAATG CATCTTTAGT 16560
 50 AATTACTTTA TCTTTAATTT CAGTCACGAT TTAATAGGCT CCTTATTTTT GATATTGATG 16620
 TCATTATAAC ACTGTCTTAA ATTTCCATGA AAAATAGTCT TAAGACGATG AGTCATGATA 16680

55

	CATCATTTTA	ACAATATCTT	TAAAAGCAGC	ATGTGGAATG	GCTAAATCTT	CTAAATCTGC	16800
	CATAGAAAAT	TCAAGATTGA	TATCATGTGG	TCGCTGTTCA	GCAAGTTTAT	GCACAAAGTC	16860
5	AGGTTCTGTG	ACAAAAGGCG	AAGACATGCC	GACCATATCT	GCATGTTGTA	AAGCATCTAA	16920
	AGCAGACTCT	GGAGAATTAA	TCCCGCCACT	TGCAATTAAA	GGGATACGAC	CTGCTAAATG	16980
10	TTCATAGACA	ATTTGGTTAA	CTGGTCGACC	GAAATGATCA	CCTGGTGTA	GAGACGTATT	17040
	TTGATAAATA	TGTCGACCCC	AGCTAGCGAT	TGCTAAGTAT	TGGATGTTTG	AAACGTCCAT	17100
	GACCCAATTG	ATTAATTGGT	TGAACTCGTC	AATGGTATAT	CCTAAATCAC	TGCCTCTGGT	17160
15	TTCTTCTGGC	GTTGCTCGAA	ATCCTAAAAT	AAAATTGTCA	GGTGCTTCTT	TATCAATCAC	17220
	TTCTTGTA	GCACGCATAA	CTTCTAAACA	TAATCTTGCA	CGATTTTTTA	ATGAGTCGGC	17280
	ACCGTAATGG	TCTGTACGTT	TATTCGAAAA	AGTTGAGAAA	AATGTTTGAA	TCAGCAAACG	17340
20	TTGTGCAATC	GAAATTTCCA	CACCATCAAA	ACCTGCTTTA	ATCGCGCGTA	ATGTAGCATC	17400
	GCGATACTGC	TGAATGATGC	TATTGATTTT	CTCATGAGAC	ATGGCGATAA	CATCGTGTTC	17460
	AATCGGTGAA	TGCAATGTCA	TAGGGCTTGG	TCCATACACC	TTTCCAAAAT	TTAAAATGGC	17520
25	TTGATTTGAA	AAACGACCAG	CATGCGCTAG	CTGGATAATA	GCGAGGCTAC	CATGTTGTTT	17580
	CATCGTAGAT	GCCATGTTAG	TTAATCCAGG	GATACAAGCA	TCATGATCAA	TATTAAAGCC	17640
	ATATTCAAAC	AATTGACCAT	AAGGTTCAAT	GTAAGCAGCG	CCGGTGACTT	GCATTCCAGC	17700
30	TGAATTAGAG	CGACGTGCAG	CATAAGCCAA	GTCTTCTTTT	GTAATATAGC	CTTCTTTTGT	17760
	TGATGTGTTT	ACGGTCATTG	GTGATAATAC	AAAGCGATTG	GAAATTTTGA	TGCCATTAGG	17820
	TAAGTGGATT	GATTGTAAAA	GTGGTTTGTA	TCGGTACATA	CTATGATTCC	TTTTCTATTG	17880
35	AATATTGTTT	TCAAAGTACC	ATGGAAAGAA	TGAATAATCA	ATGATGAACA	GTCTTGATAG	17940
	AATAGAATTG	GTACATGGAA	AGTATTTTAA	AAATTAAACT	AATGAATGGC	ATTTGTAGGT	18000
40	CTGAAAATAT	GAATATGAAA	AAGAAAAATA	AAGGCGAAAA	GATATAAAAG	TTAATTGAAA	18060
	AACGTTATCA	TATACGTGGG	TATATGAAGA	GGGAATGGTA	TTAAGAACGC	TAAAATGTTA	18120
	TGTCGGTTTG	ACATGACAGG	ATAAGTTTGG	AGATGACGGA	TTGGTTAAAT	TAAGCGTATT	18180
45	AGACTATGCC	TTAATAGATG	AAGGTAAGGA	TGCACAAAAG	GCATTGCAAG	ATTCAGTGAC	18240
	ACTTGCAAAA	TTAGCAGATC	GACTTGGCTT	TAAGCGAATT	TGGTTTACGG	AACATCATAA	18300
	TGTACCAGCG	TTTGCGTGTA	GTAAGTCCAGA	ACTTTTGATG	ATGCATACAT	TGGCGCAGAC	18360
50	AAATCACATA	CGAGTTGGCT	CTGGTGGTGT	GATGCTGCCG	CACTATCGAC	CTTATAAAAT	18420
	TGCTGAGCAT	TTTAGAATGA	TGGCAGCGTT	ATATCCAAAT	CGTATTGATT	TAGGTATTGG	18480

55

	TAGTTACGAT GAATCGATTT CGTTATTACG TGATTATCTT ACAATAAAGG ATAAACCAAG	18600
	TGCGCATACG TTAGGTGTCC AACACACAT TGATCATTTT CCAGAAATGT GGTTATTAAG	18660
5	TAGTAGCGCA ACATCTGCCA AAATAGCTGC CGAACTAGGT ATAGGGCTTT CTGTTGGAAC	18720
	ATTTTGTCTA CCAGATATAA ATGCGATACA TACAGCGAAG GATAACATTG ATATTTACAA	18780
	AAAACATTTT CAAGCATCAA CGATTAAAAT GGACGCAAAG GTGATGGCAT CTGTATTGTG	18840
10	CATTGTAGCT GATAACGAAG CGGAAGTAGC AGCATTACAA CATGCCTTAG ATGTTTGGTT	18900
	ATTAGGTAAA TTACAATTTG CAGAATTTGA AGATTTTCCT TCAGTAGACA CAGCACAAAA	18960
	GTATAAGCTT AATGATCGAG ACAAAGAGAT GATTCAAGCA CATCAAGCAC GCATCATTCG	19020
15	AGGTACACAA GAAAAGGTTA AAGCACAATT AGATGATTTC ATTGCTACGT TTGAAGTTGA	19080
	TGAGGTGTTA GTAGACCGC TTATTCCAGG TATTGAACAG CGTTGTAAAA CATTAAAATT	19140
20	ACTCGCGGAA ATTTATTTGT AGCATTTTAA ATAGAAGAGA AAGGATGAAG ATAAGATGAA	19200
	AAAGTTAGCC AATTATTTAT GGGTAGAAAA AGTAGGAGAT TTGTATGTGT TTAGTATGAC	19260
	ACCTGAATTG CAAGATGATA TTGGGACAGT AGGTTATGTT GAATTCGTAA GTCCAGATGA	19320
25	AGTTAAAGTG GATGATGAAA TTGTGAGTAT CGAAGCATCG AAAACGGTCA TTGATGTGCA	19380
	AACGCCATTG TCAGGAACGA TTATTGAGCG AAATACAAAA GCGGAAGAAG AACCGACAAT	19440
	TTTAACTCT GAAAAACCAG AAGAAAATTG GTTGTTCAAA TTGGATGATG TCGATAAAGA	19500
30	AGCATTCCTA GCATTACCGG AGGCTTAAAT GGAAACGTTA AAATCAAATA AAGCGAGACT	19560
	TGAATATTTA ATCAATGATA TGCATCGAGA GAGAAATGAC AATGACGTAT TGGTAATGCC	19620
	ATCTTCATTT GAAGATTTGT GGAATTATA TCGAGGCTTA GCAAATGTCA GACCGGCATT	19680
35	ACCTGTAAGT GATGAATATT TAGCTGTACA AGATGCTATG TTAAGTGATT TGAATCGTCA	19740
	ACATGTTACG GATTTGAAGG ATTTGAAGCC GATAAAAGGT GACAATATCT TTGTTTGGCA	19800
	AGGTGATATC ACGACGTTAA AAATCGATGC TATTGTAAAT GCTGCAAATA GTCGTTTTCT	19860
40	AGGATGTATG CAAGCTAATC ATGACTGCAT TGATAATATT ATTCATACAA AAGCGGTGT	19920
	TCAAGTTCGA CTTGATTGTG CAGAGATCAT TCGACAACAA GGGCGCAATG AAGGTGTAGG	19980
45	TAAAGCCAAA ATAACACGTG GATATAATTT GCCAGCAAAG TATATAATTC ATACGGTTGG	20040
	TCCGCAAATA CGTCGATTGC CTGTTTCAAA GATGAATCAG GACTTGTTAG CTAAATGTTA	20100
	TCTTAGCTGT CTAAATTTGG CTGATCAACA TAGTTTAAAT CATGTCGCTT TTTGCTGTAT	20160
50	ATCTACAGGT GTATTTGCTT TTCCTCAAGA TGAAGCAGCA GAAATGCTG TTCGAACAGT	20220
	AGAAAGCTAT CTCAAAGAAA CAAATTCAAC ATTGAAAGTC GTGTTCAATG TATTTACAGA	20280

55

	CAATGTCTCT GTTAATGGAT GACAAGACAA AGCAGGCTGA AGTATTGCGT ACTGCGATTG	20400
	ATGAAGCAGA TGCGATAGTG ATTGGAATTG GTGCAGGCAT GTCTGCATCT GACGGATTTA	20460
5	CATATGTAGG AGAGCGTTTT ACGGAAAATT TCCCAGATTT TATTGAAAAA TATCGCTTCT	20520
	TTGATATGTT GCAAGCGAGT TTACATCCTT ATGGCAGTTG GCAAGAGTAT TGGGCATTTG	20580
	AGAGTCGTTT TATTACATTA AACTATTTAG ATCAACCTGT AGGTCAGTCT TACCTCGCTT	20640
10	TAAAAACCTT GGTGGAAGGT AAACAGTACC ACATTATAAC TACGAATGCA GATAATGCTT	20700
	TCGATGTAGC TGATTATGAT ATGACTCATG TATTTTCATAT ACAAGGGGAG TATATACTGC	20760
	AACAGTGTAG CTCAGCATTG TCATGCTCAA ACGTATCGCA ATGATGATTT AATTCGTAAA	20820
15	ATGGTTGTTG CGCAACAAGA TATGCTTATA CCTTGGGAGA TGATTCCAAG ATGTCCAAAA	20880
	TGTGATGCCC CAATGGAAGT GAATAAACGT AAAGCGGAAG TTGGGATGGT TGAAGATGCT	20940
20	GAATTTTCATG CGCAACTACA TCGTTATAAT GCTTTTCTAG AGCAACATCA AGATGATAAA	21000
	GTGTTGTATT TGGAAATTGG AATTGGTTAT ACTACACCAC AATTGTGTA GCATCCTTTT	21060
	CAGCGTATGA CACGTAAAAA TGAAAATGCC CTTTATATGA CGATGAATAA AAAGGCATAT	21120
25	CGCATTCGGA ATTCAATTCA AGAACGTACC ATACATTTAA CTGAGGATAT CTCAACATTG	21180
	ATTACAGCAG CACTCCGGAA CGACAGCACA ACGAAAAATA ACAACATTGG AGAGACAGAA	21240
	GATGTACTTA ATAGAACCGA TTAGAAATGG AGAATATATT ACTGATGGTG CGATTGCACT	21300
30	CGCTATGCAA GTTTATGTTA ACCAGCATAT CTTTTTAGAT GAAGATATTT TATTCCCTTA	21360
	TTATTGTGAT CAAAAGTGG AAATTGGACG TTTTCAAAT ACTGCTATAG AAGTGAATCA	21420
	AGATTATATA GATAAACACA GTATTCAAGT AGTTCGCCGA GATACTGGTG GTGGCGCTGT	21480
35	GTATGTTGAT AAAGGTGCCG TTAATATGTG TTGTATTTTA GAACAAGACA CTTCAATTTA	21540
	TGGTGATTTT CAACGATTTT ATCAACCAGC TATAAAGGCG TTGCATACAT TAGGTGCAAC	21600
40	AGATGTGGTA CAAAGCGGTA GAAATGATTT AACATTGAAT GGTAAAAAAG TGTCAAGCGC	21660
	CGCAATGACA TTAATGAATA ATCGTATTTA TGGCGGTTAT TCGCTATTAC TTGATGTTAA	21720
	TTATGAAGCA ATGGATAAAG TGTTAAAGCC TAATCGCAAA AAGATTGCAT CGAAAGGGAT	21780
45	TAAATCTGTG CGCGCACGTG TTGGTCATCT TAGAGAAGCA CTGGATGAAA AGTATCGTGA	21840
	TATAACCATT GAAGAATTTA AAAATTTAAT GGTGACGCAG ATTTTGGGAA TCGATGACAT	21900
	TAAAGAGGCG AAACGATATG AATTAACGGA TGCAGATTGG GAAGCGATTG ATGAATTAGC	21960
50	TGATAAAAAG TATAAAAATT GGGATTGGAA TTATGGCAAG TCACCCAAAT ATGAATACAA	22020
	TCGAAGTGAA AGATTATCTT CAGGTACGGT AGACATAACA ATTTCTGTTG AACAAAATCG	22080

55

5 AGAAGCATT A CAAGGAACAA AAATGACAAG AGAAGATT A ACGCATCAGT TAAAGCAATT 22200
 AGACATCGTT TATTATTTTG GCAATGTTAC GGTAGAAGCA TTAGTGGATA TGATTTTAAG 22260
 10 TTAATATTGT TATTTTATGT ATGCTGAATC ATTGGAAGTG TTTGCTTGCT CTTGAAAAGG 22320
 TGACAATAGT GTTTGGTGAA GGTGAACAT ATGAGTGGAA ATTATTGCCT TTAAGTATTC 22380
 AAAGTATGAT ATATATATGG TTTTGTTC TAAATGATTG GGTATTGAA AATAGATGAG 22440
 15 TTTAATATTT TAAGGAATAT AATGATGTTT ACTTTTATAA TTCATATAGA ATATTAAGCA 22500
 ATATAAGTCT GTTGATATAT ACAAATATA ATGACTGCTA TAATGAGTAA TCAATAGACA 22560
 CAAAGAGGAG ATTATGTGAT GAATAATAAA GTATTAGTAA CCGGTGGTAC AGGGTTTGTT 22620
 20 GGCATGCGAA TTATTTACAG ATTATTAGAA CAAGGTTATG ACGTACAAAC GACGATACGT 22680
 GATTTAAGTA AAGCTGATAA AGTAATTAAA ACAATGCAAG ACAATGGCAT TTCCACAGAG 22740
 CGATTAATGT TTGTCGAAGC GGATTTATCA CAAGATGAAC ATTGGGATGA AGCAATGAAA 22800
 GATTGCAAGT ATGTCTTGAG TGTAGCATCT CCGGTGTTTT TCGGTAAAAC AGACGATGCA 22860
 GAAGTGATGG CGAaCTGcAA TTGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC 22920
 25 GGGTGTTAAA CGTGTGGTAA TGA CTGCAAA CTTTGGTGCA GTTGGTTTTA GTAATAAAGA 22980
 TAAAAATTCA ATCACAATG AAAGTCATTG GACAAATGAA GATGAACCAG GCTTATCAGT 23040
 ATATGAAAA TCAAAATTGT TAGCTGAAAA GGCAGCGTGG GATTTTGTG AGAATGAAAA 23100
 30 TACAACAGTA GAATTTGCCA CAATCAATCC AGTTGCAATT TTTGGGCCAT CATTAGATGC 23160
 ACACGTTTCA GGAAGCTTTC ATTTATTAGA AAATTTATTG AATGGTTCAA TGAAACGTGT 23220
 ACCGCAAATT CCGTTAAATG TTGTTGATGT GAGAGACGTA GCTGAACTGC ACATTTTGGC 23280
 35 AATGACAAAT GAACAAGCTA ATGGCAAGCG ATTTATTGCG ACGGCTGATG GACmAAATTwa 23340
 tTTGTTGGGA ATTGcCAAAt TAATTAAAGA AAAGGGCCTG GAAATAGCTC CAAAAGTTCC 23400
 40 TACTAAAAA TTACCCAGCT TTATTTTGAG CnAnGnGCC 23439

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4522 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCCTTTGAGA GTATATCATC TAGTCAAATT ATGCCTGTCA TTAGAGCGAC TAGCTTTGAT 60

	TATTATGCAG TCGATTTAGG GAAATCATAT CGTCTAATTG ACGAAAGCAT GTTAGAGGAT	180
	TTGAAGTTAA CTGAACAACA AATAAGAGAA ATGTCTCTGT TTAATGTTAG AAAATTGTCA	240
5	AATTCATATA CGACTGATGA AGTAAAAGGT AATATTTTTT ATTTTATTAA CTCAAATGAC	300
	GGGTATGATG CAAGTAGGAT ACTAAATACT GCATTTTTTAA ATGAAATTGA GGCACAATGT	360
	CAAGGCGAAA TGCTCGTAGC AGTGCCACAC CAAGATGTGT TAATTATTGC AGATATACGC	420
10	AATAAACAG GATATGATGT GATGGCACAT TTAACAATGG AATTTTTCAC TAAAGGTCTA	480
	GTTCCAATTA CATCATTATC CTTTGGATAT AAACAGGGTC ATCTTGAACC GATATTTATT	540
	TTAGGTAAAA ATAATAAACA AAAAAGAGAT CCAAACGTGA TTCAGCGTTT AGAAGCAAAT	600
15	CGTCGTAAAT TTAATAAAGA TAAATAGAAA TAATTGGATA AGGAGTTTTG TCATAATGAA	660
	TTTATTTTAC AATCCTAAAT ATGTAGGAGA TGTCGCATTT TTACAAATTG AACCAGTTGA	720
	AGGTGAATTA AACTACAATA AAAAAGGTAA TGTTGTTGAA ATTACTAATG AAGGTAATGT	780
20	TGTAGGTTAT AATATTTTTG AAATTTCAAA AGATATAACA ATTGAAGAAA AAGGTCATAT	840
	TAAATTAECT GATGAACTTG TAAATGTATT CCAAAGCGT ATTTCAGAAG CTGGTTTTGA	900
25	TTATAAATTA AATGCTGATC TATCACCGAA ATTTGTAGTT GGCTACGTTG AACTAAAGA	960
	CAAACATCCT GATGCAGATA AATTAAGTGT ACTAAATGTA AACGTTGGAA ATGACACATT	1020
	ACAAATTGTA TGTGGCGCGC CTAACGTTGA AGCTGGACAG AAAGTTGTTG TTGCTAAAGT	1080
30	AGGTGCAGTG ATGCCTAGCG GTATGGTAAT TAAAGATGCT GAATTACGTG GTGTTGCCTC	1140
	AAGCGGTATG ATTTGTTCAA TGAAAGAATT GAATTTACCT AATGCACCTG AAGAAAAAGG	1200
	TATTATGGTA TTAAATGACA GCTATGAAAT TGGACAAGCA TTtTTTGAAT AATTAAGGAA	1260
35	GGTAGTGAAA ATATGAGCTG GTTTGATAAA TTATTCGGCG AAGATAATGA TTCAAATGAT	1320
	GACTTGATTC ATAGAAAGAA AAAAAGACGT CAAGAATCAC AAAATATAGA TrACGATCAT	1380
	GACTCATTAC TGCCTCAAAA TAATGATATT TATAGTCGTC CGAGGGGAAA ATTCCGTTTT	1440
40	CCTATGAGCG TAGCTTATGA AAATGAAAAT GTTGAACAAT CTGCAGATAC TATTTTCAGAT	1500
	GAAAAAGAAC AATACCATCG AGACTATCGC AAACAAAGCC ACGATTCTCG TTCACAAAAA	1560
45	CGACATCGCC GTAGAAGAAA TCAACAACCT GAAGAACAAA ATTATAGTGA ACAACGTGGG	1620
	AATTCTAAAA TATCACAGCA AAGTATAAAA TATAAAGATC ATTCACATTA CCATACGAAT	1680
	AAGCCAGGTA CATATGTTTC TGCAATTAAT GGTATTGAGA AGGAAACGCA CAAGCCAAAA	1740
50	ACACATAATA TGTATTCTAA TAATACAAAT CATCGTGCTA AAGATTCAAC TCCAGATTAT	1800
	CACAAAGAAA GTTTCAAGAC TTCAGAGGTA CCGTCAGCTA TTTTGGCAC AATGAAACCT	1860

55

	AAACAAAAAT ATGATAAATA TGTAAGCTAAG ACGCAAACGT CTCAAAATAA ACAATTAGAA	1980
	CAAGAAAAAC AAAATGATAG TGTGTGCAAA CAAGGAACTG CATCTAAATC ATCTGATGAA	2040
5	AATGTATCAT CAACAACAAA ATCAATGCCT AATTATTCAA AAGTTGATAA TACTATCAAA	2100
	ATTGAAAATA TTTATGCTTC ACAAATTGTT GAAGAAATTA GACGTGAACG AGAACGTAAA	2160
	GTGCTTCAAA AGCGTCGATT TAAAAAAGCG TTGCAACAAA AGCGTGAAGA ACATAAAAAAC	2220
10	GAAGAGCAAG ATGCAATACA ACGTGCAATT GATGAAATGT ATGCTAAACA AGcGGAACgC	2280
	TATGTTGGTG ATAGTTTATT AAATGATGAT AGTGACTTAA CAGATAATAG TACAGATGCT	2340
	AGTCAGCTTC ATACAAATGG CATAGAGAAT GAAACTGTAT CAAATGATGA AAATAAACAA	2400
15	GCGTCAATAC AAAATGAAGA CACTAATGAC ACTCATGTAG ATGAAAGTCC ATACAATTAT	2460
	GAGGAAGTTA GTTTGAaTCA AGTATCGACA ACAAACAAT TGTCAGATGA TGAAGTTACG	2520
	GTTTCGAATG TAACGTCTCA ACATCAATCA GCACTACAAC ATAACGTTGA AGTAAATGAT	2580
20	AAAGATGAAC TAAAAAATCA ATCCAGATTA ATTGCTGATT CAGAAGAAGA TGGAGCAACG	2640
	aATAAAGAAG AATATTCAGk AAGTCAAATC GATGATGCAG AATTTTATGA ATTAAATGAT	2700
	ACAGAAGTAG ATGAGGATAC TACTTCAAAT ATCGAAGATA ATACCAATAG AAACGCGTCT	2760
25	GAAATGCATG TAGACGCTCC TAAACGCAA GAGTACGCAG TAACTGAATC TCAAGTAAAT	2820
	AATATCGATA AAACGGTTGA TAATGAAATT GAATTAGCAC CGCGTCATAA AAAAGATGAC	2880
30	CAAACAAACT TAAGTGTCAA CTCATTGAAA ACGAATGATG TGAATGATAA TCATGTTGTG	2940
	GAAGATTCAA GCATGAATGA AATAGAAAAG AATAACGCAG AAATTACAGA AAATGTGCAA	3000
	AACGAAGCAG CTGAAAGTGA ACAAATGTC GAAGAGAAAA CTATTGAAAA CGTAAATCCA	3060
35	AAGAAACAGA CTGAAAAGGT TTCAACTTTA AGTAAAAGAC CATTTAATGT TGTCATGACG	3120
	CCATCTGATA AAAAGCGTAT GATGGATCGT AAAAAGCATT CAAAAGTCAA TGTGCCTGAA	3180
	TTAAAGCCTG TACAAAGTAA GCAAGCTGTG AGTGAAAGAA TGCCTGCGAG TCAAGCCACA	3240
40	CCATCATCAA GATCTGATTC ACAAGAGTCA AATACAAATG CATATAAAC AAATAATATG	3300
	ACATCAAACA ATGTTGaGAA CAATCAACTT ATTGGTCATG CAGAAACAGA AAATGATTAT	3360
	CAAAATGCAC AACAATATTC AGAGCAGAAA CCTTCTGTTG aTTCAACTCA AACGGAAATA	3420
45	TTTGAAGAAA GTCAAGATGA TAATCAATTG GAAAATGAGC AAGTTGATCA ATCAACTTCG	3480
	TCTTCAGTTT CAGAAGTAAG CGACATAACT GAAGAAAGCG AAGAAACAAC ACATCCAAAC	3540
50	AATACTAGTG GACAACAAGA TAATGATGAT CAACAAAAAG ATTTACAGTC ATCATTTTCA	3600
	AATAAAAATG AAGATACAGC TAATGAAAAT AGACCTCGGA CGAACCAACA AGATGTTGCA	3660

55

	CCAAGTGTTT CATTACTAGA AGAACCACAA GTTATTGAGT CGGACGAGGA CTGGATTACA	3780
	GATAAAAAGA AAGAACTGAA TGACGCATTA TTTTACTTTA ATGTACCTGC AGAAGTACAA	3840
5	GATGTAAC TG AAGGTCCAAG TGTTACAAGA TTTGAATTAT CAGTTGAAAA AGGTGTTAAA	3900
	GTTTCAAGAA TTACGGCATT ACAAGATGAC ATTAAAATGG CATTGGCAGC GAAAGATATT	3960
	CGTATAGAAG CGCCTATTCC AGGAACTAGT CGTGTTGGTA TTGAAGTTCC GAACCAAAAT	4020
10	CCAACGACAG TCAACTTACG TTCTATTATT GAATCTCCaA GTTTTAAAAA TGCTGAATCT	4080
	AAATTAACAG TTGCGATGGG GTATAGAATT AATAATGAAC CATTACTTAT GGATATTGCT	4140
	AAAACGCCAC ACGCACTAAT TGCAGGTGCA ACTGGATCAG GGAAATCAGT TTGTATCAAT	4200
15	AGTATTTTGA TGTCTTTACT ATATAAAAAT CATCCTGAGG AATTAAGATT ATTACTTATC	4260
	GATCCAAAAA TGGTTGAATT AGCTCCTTAT AATGGTTTGC CACATTTAGT TGCACCGGTA	4320
20	ATTACAGATG TCAAAGCAGC TACACAGAGT TTAAATGGG CCGTAGAAGA AATGGAACGA	4380
	CGTTATAAGT TATTTCACACA TTACCCATGT ACGTAnTATA ACAGCATTTA ACnAAAAAGC	4440
	CCCATATGAT GAAAGAATGn CAAAAATTGT CATTGTaATT GATGAGTTGG CTGATTTAAT	4500
25	GATGATGGTC CGCAAGAAGT TG	4522

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 751 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

	TCAAGTTTAC GGATACGTAT ATATTTTGCA TGACATTAG TGCAATAATA TTCATAATTT	60
	GCCCGTTGTT GATAGCTTTC AATGCTGTTA CAAAATCTAG GCGCTCCAAC CTGTTGGCTC	120
40	AATCGTTTAA AATCTTGATC TTTATGTTGA TAACCTTTAC CAGCAATATG CAAGTGATAA	180
	TGACACAATT CGTGCAGTAT AATTTTTTACA ACAGCATCTT CTCCATAATG CTCATATTGT	240
	TTTGGATTAA TTTCAATATC ATGGGACTTT AAAAGATAAC GTCCGCCTGT TGTACGTAAC	300
45	CTTTTATTAA AATATGCACA ATGTCGAAAC GTACGTCCAA ATTTTCTTC CGAAAGATTC	360
	TCAACCATTG GCTGAAGTTT GTCATTATTC ATGTGGATCA ATCATCGTTA ATGATACTTT	420
50	GTCTTTATTT TTGTCAATAC TGTAATCCA AACGTCAACG ATATCACCAA CACTGACAAT	480
	ATCCATTGGA TTTTTCACGA ACTTCTTAGA AAGTTTCGAA ACATGGACAA GTCCATCTTG	540

TTTCATTCTTCT TCTTGTAAT CTTCAATTGA TAGCACATCG GATTTAAGGA TTGGTGTTC 660
 AAACTCGTCC CTTGGATCTC GATTAGGTGC GTTCAAGGAT TTAATAATAT CCTCTAATGT 720
 5 AGGTACACCG ACTTGTAATT CAATCGCCAG T 751

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1076 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TCTCCAGCTT TAACCTTGATC TGGCACTTTA ACAATTGTCT GATCCATACA TACGCGACCA 60
 20 ATAACTTCGC ATTGATGACC ATTTACATTT ACAAAGCTAC CTTGCATTAT GCGTAAATGG 120
 CCATCTGCAT ATCCAATAgG TAACAATGCT ATTGTAGTTG GGTCACTAGC TGTATAAGTT 180
 GCACCATAAC TTACAGACTC ACCCGCTTGT AGCGTCTTTG TTTGAACTAC ATTAGCAATT 240
 25 AATTGCACAC TTGGTTTAAG GTGTACTTTA ACTTTTTGCT GTACATACTC TGATGGATAA 300
 TATCCATAAA GGGAAATTC TGGTCTTATT GCATTACAGA ATTGGCAATC CATTAAATAGA 360
 GAGCCTGCTG AGTTCTGACA ATGTATATAT TCAGGTTTAA TTGCTTCATT GACCATATCT 420
 30 TTAACACGTT GATATTGTTC AGTTGTCATA TCTCCTGGTT CGTCAGCACA GGCAAAGTGT 480
 GTAAACACGC CTTCAAATAC AAGTTGCTCA TATTGTTGAA TGATTTCAAT CACTTCTTGA 540
 TACGTTTTAG TATCTTTAAT ACCTAAACGT CCCATTCTCTG TATCTAATTT AATGTGCAAC 600
 35 CATAACTTTT TCTCTTGCTC ACCAGAAATG TTTTAAATG CTTCTTTCAA CCACTGTTTA 660
 GACGGAACCG TTAAGGCAAC TCGGTGTTGT ATCGCTTTAT CAATATCTTT AGCTGGTAAC 720
 ACACCTAAGA CTAAAATTTT AGCAGTAATC CCATGCATTC TAAGTTCTAT CGCTTCATCT 780
 40 AACGTTGCTA CAGCAAAAAA TGTGGCGCCA TTTTCCATTA AATGACGTGC TACTTTAACA 840
 CTACCTAGTC CATAGGCATT GGCTTTAACG ACAGCCATCA CTGTTTTATT TGGATGCAAT 900
 GTACTGAATA CTTTGAAATT TGATGCAACA GCGTTTAAAT CTACATTCAT ATACGCAGAT 960
 45 CTATAATATT TATCCGACAT ATTACTTCCT CCTGTAATTC CCACACGTTT TAAAACTAGA 1020
 TCTTAATTAT CATTGTATAA CAAATTTAAA ATGCTGACTT TTCTAAAACA ACTTGG 1076

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2930 base pairs

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	TGACCACAAT GCCCAATACA ACCATCCCAT GGTAAAGCCA AGAGATGAGT CAATAAAGCG	60
10	TGTTGAATAA GAGCTGAATG AACCTGATAC TGGATAAAAT GTTGCCAACT CTCCAATTGA	120
	TGACATTAAG AAATATAGCA TGACACCAAT AACAAGATAA GCGAGTATAG CGCCTCCAGG	180
	ACCAGCTTGA GAAATGATAT TACCAGTAGC TACAAATAGA CCAGTCCCAA TTGCACCACC	240
15	TATAGCAATC ATGGAAATGT GTCTTGAGTT AAGACTACGG TTCATTTTAT TATCTTCCAT	300
	ATTTAGTCTC CCATCTATTT AAATATACCC ATTATTGTAA GCTTTTTAAG TGTACTATTC	360
	AATAACTATT TAGTACTGTA AAGCGAAAAA ATTAAAATTT TCTGATTTT TAATCATCTT	420
20	GAGCATGTTT AATTGTAATT TTGATGGGGT TAAATTATAA TATGTATTAA ATTATAATTA	480
	TnATAAATTG TGGAGGGaTG ACTATGTCAC AACAAGACAA AAAGTTAACT GGTGTTTTTG	540
	GGCATCCAGT ATCAGACCGA GAAAATAGTA TGACAGCAGG GCCTAGGGGA CCTCTTTTAA	600
25	TGCAAGATAT TTACTTTTTA GAGCAAATGT CTCAATTTGA TAGAGAAGTA ATACCAGAAC	660
	GTGCAATGCA TGCCAAAGGT TCTGGTGCAT TTGGGACATT TACTGTAACT AAAGATATAA	720
	CAAAATATAC GAATGCTAAA AaATTCTCTG AAATAGGTAA GCAAACCGAA ATGTTTGCCC	780
30	GTTTCTCTAC TGTAGCAGGA GAACGTGGTG CTGCTGATGC GGAcGTGACA TTCGAGGATT	840
	TGCGTTAAAG TTCTACACTG AAGAAGGGAA CTGGGaTTTA GTAGGGAATA ACACACCgGT	900
	ATTCTTCTTT AGAGATCCAA AGTTATTTGT TAGTTTAAAT CGTGCGGTGA AACGAGATCC	960
35	TAGAACAAAT ATGAGAGATG CACAAAATAA CTGGGATTTT TGGaCGGGTt TCCAGAAGCA	1020
	TTGCACCAAG TAACGATCTT AATGTCAGAT AGAGGGATTC CTAAAGATTT ACGTCATATG	1080
40	CATGGGTTTCG GTTCTCACAC ATACTCTATG TATAATGATT CTGGTGAACG TGTTTGGGTT	1140
	AAATTCCATT TTAGAACGCA ACAAGGTATT GAAAACTTAA CTGATGAAGA AGCTGCTGAA	1200
	ATTATAGCTA CAGATCGTGA TTCATCTCAA CGCGATTTAT TCGAAGCCAT TGAAAAAGGT	1260
45	GATTATCCAA AATGGACAAT GTATATTCAA GTAATGACTG AGGAACAAGC TAAAAACCAT	1320
	AAAGATAATC CATTTGATTT AACAAAAGTA TGGTATCACG ATGAGTATCC TCTAATTGAA	1380
	GTTGGAGAGT TTGAATTAAA TAGAAATCCA GATAATTACT TTATGGATGT TGAACAAGCT	1440
50	GCGTTTGAC CAACTAATAT TATTCCAGGA TTAGATTTTT CTCCAGACAA AATGCTGCAA	1500
	GGGCGTTTAT TCTCATATGG CGATGCGCAA AGATATCGAT TAGGAGTTAA TCATTGGCAG	1560

55

GGTCAAATGC GCGTAGTTGA CAATAACCAA GGTGGAGGAA CACATTATTA TCCAAATAAC 1680
CATGGTAAAT TTGATTCTCA ACCTGAATAT AAAAGCCAC CATTCCCAAC TGATGGATAC 1740
5 GGCTATGAAT ATAATCAACG TCAAGATGAT GATAATTATT TTGAACAACC AGGTAAATTG 1800
TTTAGATTAC AATCAGAGGA CGCTAAAGAA AGAATTTTTA CAAATACAGC AAATGCAATG 1860
GAAGGCGTAA CGGATGATGT TAAACGACGT CATATTCGTC ATTGTTACAA AGCTGACCCA 1920
10 GAATATGGTA AAGGTGTTGC AAAAGCATTG GGTATTGATA TAAATTCTAT TGATCTTGAA 1980
ACTGAAAATG ATGAAACATA CGAAAACTTT GAAAATAAA TTTGATATGT AGTTTCTATA 2040
TTGCGTAGTT GAGCAGTTTA TGATATCATA ATAAATCGTA AAGATTCCTA ACAAGAGAGG 2100
15 GTGTTTAACG TGCGCGTAAA CGTAACATTA GCATGCACAG AATGTGGCGA TCGTAACTAT 2160
ATCACTACTA AAAATAAACG TAATAATCCT GAGCGTATTG AAATGAAAAA ATATTGCCCA 2220
AGATTAAACA AATATACGTT ACATCGTGAA ACTAAGTAAT TCTTATCATT CAAATACGAC 2280
20 GATTTGAAAA TAAAGCGGGC TTACCTATTA TATTGGGGAG CTCGCTTTTT TATGAAATTT 2340
TTGTGAAGAG TGATTAATGG ATTGAGTTTC ATCGGTAGAA CAATATATGA TTATATTAGT 2400
TGTTACTTTA TTAaaaTTTG AGAATATTTA TAGAAGGAAA TAGATTACTG ATTTTATAAA 2460
25 GTCACTTTGT TAGCGAATGC TTGAAAGAGT ATTTAATATA GTAGAATTTA AAATTTCAAA 2520
GCGGAATTTA ATAAGTACGA AGTAGTTCTG GGTATGTTTT ATAAATGTTT GATAATACAC 2580
TTTAATCTTA AATATGATGG TTTAGAAAAT GATTTAACAA AGAAATGAaa CTTTACTGTT 2640
30 GAATTATGTG AGGATTGTGT TATTATATAA ATCGTAATAA TTACGATTTG ATAAAAAGTG 2700
AGGTAACAT ATATGGCTAA GAAATCTAAA ATAGCAAAAG AGAGAAAAAG AGAAGAGTTA 2760
35 GTAAATAAAT ATTACGAATT ACGTAAAGAG TTAAGGCAA AAGGTGATTA CGAAGCGTTA 2820
AGAAAATTAC CAAGAGATTC ATCACCTACA CGTTTAATA GAAGATGTAA AGTAACTGGA 2880
AGACCTAGAG GTGTATTACG TAAATTTGAA ATGTCTCGTA TTGCGTTTAG 2930

40 (2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3606 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

50 CTTCTTGCCA TGGCTCTCTT TATTTAAAAA TGCTTCCAAC TTGTCCATTT GATTGTTTCT 60

55

	TTATAAAAAA CTAATTTTAC AAATGCTTTT GCGTTCTTAC AAAAAATGCA TTTGACTATT	180
	ATTATAATAA GCGTATAATT GTCGCATATT ATTTTTTGTA TTTTGGCAA TAACGAAGGA	240
5	GTATTTATGA ATAAAGACAA GCAATTGCAC AACGACAAAA TCAATCTATC CCAATTAGTC	300
	TTATTAGGGT TAGGCTCTTT AATAGGATCT GGTTGGCTAT TTGGTGGCTG GGAAGCATCA	360
	TCAATAGCTG GACCAGCAGC AATCATATCA TGGGTTCTTG GATTCTAGT CATTGGAACC	420
10	ATTGCCTATA ACTACATTGA AATCGGCACA ATGTTTCTCTC AATCAGGTGG CATGAGTAAC	480
	TATGCCCACT ATACACATGG CTCATTATTA GGCTTTATTG CTGCTTGGGC GAATTGGGTG	540
	TCTTTGGTGA CAATAATACC TATCGAAGCT GTGTCAGCTG TTCAATATAT GAGTTCTTGG	600
15	CCGTGGCATT GGGCGAAACC AATGAGATAT TTAATGAAA ATGGCTCTAT TAGCACATAC	660
	GGATTGCTAG CTGTATATCT CATCATTGTT ATTTTTTCAT TATTAACTA TTGGTCCGTA	720
	AAACTTTTAA CATCATTTAC GAGTTTAATT TCTGTATTTA AATTAGGCGT ACCCATGTTA	780
20	ACCATCATCA TGTTGATGCT ATCAGGATTC GACACTTCAA ATTACGGCCA TTCGGCAAGC	840
	ACATTTATGC CTTACGGAAG TGCACCGATT TTTGCTGCAA CAACAGCATC AGGGATTATT	900
	TTTTCATTCA ATTCATTCCA GACAATTATT AATATGGGTT CAGAAATTAA AAATCCTGAA	960
25	AAAAATATCG CAAGAGGCAT CGCTATCTCA CTGTCAATCA GTGCAGTGTT GTACATCATT	1020
	TTACAAAGTA CGTTTATCAC TTCTATGCCT CAATCAATGT TACAACATAG TGGATGGAAT	1080
	GGCATCAACT TCAATTCACC ATTTGCTGAT TTAGCTATCT TATTAGGAAT TAATTGGCTC	1140
30	GCAATTTTAC TATACATTGA AGCTTTTGTA TCACCATTTCG GTACTGGCGT GTCATTTGTC	1200
	GCCGTTACAG GTCGAGTTTT ACGAGCAATG GAGAAAAATG GACATATCCC TAAATTTCTT	1260
35	GGGAAGATGA ATGAAAAATA TCATATCCCA CGTGTAGCAA TCATCTTTAA TGCCATCATT	1320
	AGTATGATTA TGGTTACATT ATTTAGAGAT TGGGGTACGC TAGCAGCAGT TATTTCTACT	1380
	GCAACTTTAG TAGCCTATTT AACTGGCCCA ACGACAGTGA TTGCATTAAG AAAAAATGGGA	1440
40	CCAACAATGA CTCGTCCATT TAGAGCAAAA ATTTTAAAAG TAATGGCACC ATTATCATT	1500
	GTATTAGCTT CATTAGCTAT ATATTGGGCA ATGTGGCCAA CAACGGCTGA AGTTATTTTA	1560
	ATCATTATAC TTGGATTACC AATCTACTTC TTCTATGAAT ATCGTATGAA TTGGCGTAAT	1620
45	ACAAAGAAAC AAATTGGTGG TAGCTTATGG ATTATTGTAT ATTTAATCGT GCTATCAATA	1680
	CTGTCAATTA TAGGAAGCAA AGAATTTAAA GGCTTAAATA TGATTCATA TCCATTTGAC	1740
	TTTATCGTTA TTATTATTGT GGCATTATC TTCTATTACA TCGGTACAAC GAGTTCATTT	1800
50	GAAAGCGTCT ATTTCCGTCG CGCAACACGA ATCAATACGA AGATGCGTGA GTCACTAAAT	1860

55

	CACACACATT AACCAACCAT TGATTTCAAC ATCTTG GTTTTATT TTGAAAATCG	1980
	GTTATAAATA ACTAACATAA CAAGATGATG ATCAGGCTGG GACATAAATC AATGTTCTAT	2040
5	GCTCTACGAA gTTATATTGG CAGTAGTTGA CTGAACGAAA ATGCGCTTGT AACAAAGCTTT	2100
	TTTCGATTCT AGTCAGGGGC CCCAACACAG AGAATTTCTGA AAAGAAATTC TACAGGCAAT	2160
	GCAAGTTGGG GTGGGACGAC GATAAAGAAA TACTTTTTCT ATAGAAATTA GTATyCTTA	2220
10	TGCATGAGTT TTACTCATGT ATTCATATTT TTAAGTACAC ATTAGCTGTG GCTAATGTAT	2280
	AAGAACCACT ACATAATAAA TCATTGTGG CTCTTTATCA TTTCTGTCCC ACTCCCGTAG	2340
	AAGTACATCA TATAATGCTG AAAATGGTTT GAGTTAAAC AGATATCAAG CTCGTCTGAT	2400
15	TCAGTCACAA AATTGTCTTG TTATACTTGT CACCTATCAT CTATAGACCG TGGTATGATT	2460
	AAATTGGGGA TGATAAAGGA GGTTAATAAA TATGAAGATT AATACTACAG GTGGTCAAAT	2520
	TCATGGTATT ACACAAGATG GTTTAGATAT CTTCTTAGGC ATTCTTATG CAGAACCACC	2580
20	AGTTCATGAC AATCGCTTTA AACATTCTAC GTTAAAAACA CAATGGTCAG AGCCAATTGA	2640
	TGCAACTGAA ATACAACCCA TCCCACCGCA ACCAGACAAC AAATTAGAAG ATTTTTCTC	2700
	CTCACAATCT ACAACTTTTA CTGAACATGA AGACTGTTTA TATCTAAATA TTTGGAAACA	2760
25	ACATAATGAT CAGACGAAGA AACCTGTCAT CATTTATTTT TATGGTGTA GTTTTGAAAA	2820
	TGGTCATGGT ACAGCCGAAC TCTATCAACC GGCACATTTA GTACAAAATA ACGACATTAT	2880
	CGTTATTACA TGCAATTATC GTTTAGGCGC ATTAGGATAT TTAGACTGGT CATATTTTAA	2940
30	TAAAGATTTT CATTCCAATA ATGGCCTTTC AGATCAAATC AATGTCATAA AATGGGTGCA	3000
	TCAATTTATT GAATCCTTCG GTGGCGACGC TAATAACATT ACTTTAATGG GTCAGTCTGC	3060
	AGGCAGTATG AGCATTTTGA CTTTACTTAA AATACCTGAC ATTGAGCCAT ACTTCCATAA	3120
35	AGTCGTTCTA CTAAGTGGCG CACTACGATT AGACACCCTT GAGAGTGCAC GCAATAAAGC	3180
	ACAACATTTT CAAAAATGA TGCTCGATTA TTTAGATACA GATGATGTTA CATCATATC	3240
40	GACAAATGAT ATTCCTATGC TGATGGCGAA gCTAAAACAA TCTCGAGGAC CTTCTAAAGG	3300
	GCTTGATTTA ATATATGCGC CTATTAAAAC AGATTATATA CAAAATAATT ATCCAACAAC	3360
	GAAACCAATT TTTGCATGTT ATACAAAAGA TGAAGGCGAT ATTTATATTA CTAGTGAACA	3420
45	GAAAAAATTA TCGCCGCAAC GCTTTATCGA CATTATGGAA TTAAATGATA TTCCTTTAAA	3480
	ATACGAAGAT GTTCAGACGG CGAAGcAACA ATCTTTAGCG ATTACACATT GTTATTTCaA	3540
	ACAGCCGATG aAGCAATTTT TACmACmACT CAATATACmA GATTCCAACC GCACCAACTA	3600
50	TGGCTT	3606

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15109 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

10 GAAATTAAAA AAGCAATTGG nACAAGATGC AACAGTGTCA TTGTTTGATG AATTTGATAA 60
 AAAATTATAC ACTTACGGCG ATAAGTGGG TCGTGGTGGA GAAGTATTAT ATCAAGCATT 120
 TGGTTTGAAA ATGCAACsAG AACAACAAAA GTTAACTGCA AAAGCAGGTT GGGCTGAAGT 180
 15 GAAACAAGAA GAAATTGAAA AATATGCTGG TGATTACATT GTGAGTACAA GTGAAGGTAA 240
 ACCTACACCA GGATACGAAT CAACAAACAT GTGGaAGAAT TTGAAAGCTA CTAAAGAAGG 300
 ACATATTGTT AAAGTTGATG CTGGTACATA CTGGTACAAC GATCCTTATA CATTAGATT 360
 20 CATGCGTAAA GATTTAAAAG AmAAATTAAT TAAAGCTGCA AAATAATTCA GCTATATAAG 420
 TTAGTGAAAT GAGAGTCTGA AACATATCAA TCTTTTGATA TTGTATTAGG CTCTTATTTT 480
 TATAGCTAGA AAGTTAGATA TTTGTATTTT TTAAATAAT AAGTGCCGTT GTTATCGTTC 540
 25 AATTTAATTA ATGATAGATT AGTATTATTA TAGCTAAAGT AGTATACCTG AGAAAATAGC 600
 TCAATGTATC TCTTTATTAA TAAGTTATAT CATAATTATT TTAGTGATA CTTTATGGAA 660
 GGGATATCAG GGAATGGCTT TCAATTAAAG AAGAGGTTTA AAAGGATTAC AACAGAATGT 720
 30 TATGATTTTG TAGAAAGATA TATAACAACG TTTTATAAAA ACATAATATT GTTAATGGAA 780
 AATGAAATGT AAGGGGGATT TCGAGTGAAT AAGAAAGTTT ATTTTAACCA CGATGGTGGT 840
 GTAGATGATT TAGTATCTCT ATTTTATTA TTACAAATGG AAAACGTTCA ATTGATAGGG 900
 35 GTCAGTACAA TTGGTGCTGA TTGTATTTA GAGCCATCTT TGAGCGCATC AGTAAAAATT 960
 ATTAATCGTT TTTCAATGA AGATATTCAA GTTGCGCCAT CATATGAACG AGGAAAAAAT 1020
 CCATTTCTTA AAGAATGGCG TATGCATGCC TTTTTATGG ACGCATTGCC AATTTTAAAT 1080
 GAGCCAGTCA AACATGTTGC TTCAAATGTG AGCGACAAAG AAGCCTTTGA AGACATTATT 1140
 CAAACTTTAA AGAGACAATC AGAAAAAGTA ACATTATTAT TTACAGGCCC GCTTACAGAT 1200
 45 TTAGCAAAAG CACTACAAAA AGATTCATCT ATCGTTCAGT ATATAGAAAA ATTAGTTTGG 1260
 ATGGGTGGCA CCTTTTACC AAAAGGAAAT GTTGAAGAAC CTGAGCATGA TGGTTCTGCA 1320
 GAATGGAATG CATATTGGGA TCCAGAAGCG GTTAAAATTG TTTTGTAGAG CGATATAGAG 1380
 50 ATTGATATGG TTGCTTTAGA AAGTACGAAT CAAGTACCGC TAACGTTAGA TGTTAGACAA 1440

55

	GTACCACCAT TAACACACTT TATAACAAAT TCTACTTACT TTTTATGGGA TGTTTTAACG	1560
	ACTGCTTATA TTGGTAACAA GGACTTGGTT CATTCAATTG AGAAAAAAGT CGATGTAATA	1620
5	AGTTATGGAC CAAGTCAAGG TAAGACATTT GAGTGTAAG ATGGGCGCAA AATTAATGTC	1680
	ATAAATCATG TAGATAACAA CGCATTTTTT GATTATATAA CTGCACTTGC TAAAAAGTA	1740
	AATTAACAGC TGTGTAGAAT AATTAAGGTT TTAATTTATA TAGAACAACT TATTGTAAAC	1800
10	TTTTCATTTT TTAAGTTTAA CAATGGTGCT ATAATAATGG TCATGAAATA CGAAAGGAAG	1860
	TAAAAATGA CAACAAAACA GTTAGTATAT ACAGCTTTAA TGACAGCGAT TATCGCTATT	1920
	TTAGGATTGG TACCGGTAAT TCCACTACCA TTTTCTTCAG TACCAATTGT ACTTCAAAC	1980
15	ATTGGTATTT TCTTAGCAGG TGCGATTTTA GGACGTAAAT ATGGCACATT AAGTGTTATC	2040
	GTCITTTTAT TATTAGTAGT TGCTGGCTTG CCATTGTTAT CAGGTGGTCG CGGTGGCATC	2100
	GGTGTATTCT CAGGTCCTTC AGCAGGGTTT TTACTATTAT ATCCAGTTGT AGCATTTCATG	2160
20	ATTGGGGCGA TTCGAGATAG ATTCATCAAT GAAATTAATT TCTGGATTTT ATTCGTTGGT	2220
	ATTTTAGTTT TTGGTGTTAT AGCATTAGAT GTTATTGGTA CATTGATTAT GGGCATGATT	2280
	ATTAACATAC CATTTACGAA AGCTATTTCA ATTTCATTAG CTTATTGACC TGGTGATATA	2340
25	TTAAAAGCAA TTGTAGCAAG TTTGATTGGT ACAGCTTTAC TTAATCACTC GCAGTTTCGT	2400
	CAAATTATGG GAATAAAATA ATCATATTTA AGATAGTAAA GTAATTGAAT AAGTTGCTTT	2460
	GAAATTTATA AAAGTGAAAG GAGTAGGTGT CAATGGCTAG TATAAGTATG TCAGATATAT	2520
30	ATTGTAACGG CACTATATTT GAAAATGACG ACGAGCAGTT GATTTATTTA ACGCCTTCTT	2580
	TTCCACAACG ATACACAAGT AACACATGGA TATATAAAAA GACGCCTACC CAAGAGCGAT	2640
	GGCTGAAAGA CTTAGAACGT CAACATCAAT TACATACAAA TCAAGGTTCA AATCATTATG	2700
35	CGTTTAGTTT CCCGGAAAAT GAACAACTTG ATAATCATTG GATGGCTATG TTAAAGATA	2760
	TGAATTTTGA ACTAGGTATT ATGGAATTGT ATGCCATAGA AAGTGATGCG CTGCCAATT	2820
40	TGCCGCGTAA CTCTGACGTT GAAATTGCCA TCGTTGACGA GTCGCATATA GATGCCTATT	2880
	TAAAAGTTGC ATATCAGTTT AGTTTGCCAT TTGAAAAGA CTATGCAGAT GCACATGAAG	2940
	AAATGGTAAG GGAACATTAT CAAAAAGATG TGATTAAACG CTTAGTAGCT TATTTAAATA	3000
45	ATGAACCTAT TGGCGTTGTA GATGTCATTG AAAGTGAAAA TTACATTGAA TTAGATGGAT	3060
	TTGGTGTATT AGAACAAATTT CGGCACCAAG GAATTGGATC TACAATTCAA TCGTTGATAG	3120
	GTGAATACGC CATATCAAAA AATCACAAAC CAATCATATT AGTTGCAGAT GGTGAAGATA	3180
50	CAGCAAAAGA TATGTATGCA AAGCAAGGTT ATGTCATCA ATCGTTTTGT TATCAAATAT	3240

	TAAGCTGGTT TCGAGTAGAA ATCAACTTAC TGCTTTTAA ATTGTTTGA GCTACTTATA	3360
	CTTATAAAAA TAGTGC GTTT AAATTGTTGA TTCATGTAGA ATATCGTTCA TTATGACACA	3420
5	CTATAATGAA TATGTTATTG TTCAGAATCA ATGATACGTT CTGGATGACT GTATATATTA	3480
	AAGCCACCAT TTCGAATAAA TCCAAGTCC GTAATATTTA GGTCAATAGC TAAGGTTACA	3540
10	GCAAGCGTTG TCGGAGCTGA TTTAGATAAA ATGACGCCAA CACCAATTTT TCGCGCTTTA	3600
	ATTAAAATTT CTGATGAAAT ACGTCCACTA AAAATTAATA CTTTATCTCG GACAGTAATA	3660
	TGTCGCTGAA TACAAAATCC ATATAATTTA TCTAGAGCGT TATGTCTACC AATGTCTTGT	3720
15	CGATGTACAA AAAATGTCAA ACCATCGCTT ATAGCAGCAT TATGTAAGCC ACCTGTTTCT	3780
	TGGTAAATAT GACTTGCACT TTGTAATCGA GTCATCATGT TAATAATTTG CATTGGAGTT	3840
	AAAGTGATTT TAGACATAGA TGTTTTAGCG ATAGCAGCAT CATTTTGAAA ATAAACTCA	3900
20	CGACTCTTTC CGCAACAAGA TGCAATCATT CGTTTGTGG AATATTGAAA GCGATCGCCT	3960
	AAATCTTTAT TAAGTTCAAC ATGGGCAAAA CCTTTACTAT CATCAATCAG TACAGATTTT	4020
	AATTCATCTC GCTTTAAAT GGCACCTTCC GAAGCCAGAA ATCCAATGAC TAACTCCTCA	4080
25	AGGTTTGTG GACTGCATAT AACAGTCGCA AATTCCTCAC CATTCAACAT AATTGTAAGT	4140
	GGAAATTCTG TCACATATTG ATCTGTTGTA TTGAATAATT TTCCATCTTC ATATCTAACA	4200
	ATTGGTTGAC CTAAAGATAC ATCTTTGTTT ATTATCTAAC CCCTTTAATT AGCTTAAACT	4260
30	TTATTTTAAA GCAATTTGCT TAAAATTTTA ACATATTTGC TTAAGTTTGA AATTTGATTG	4320
	ATAAAAATTA ATAGCGAGCA ATCTGTTTGA TTAAATTTGA ATTCGAGAAT ATACATACTA	4380
35	GGGCATCAAT TAATAAATAT CAATCTTATG CAAATTTGAC AATTGTTTGA ATCAATATAT	4440
	AAACAGGCAA CGTTCCTTTT CAAATATAAT AGTAAGTGTA TAATGAAAAT GTAAATATTA	4500
	TTAAATATGG GGGTTCACTC AATGAAATTG AAACGTTTAT TTGCTGTTGT GATTGCAATG	4560
40	CTTTTAGTAT TAGCTGGTTG CTCTAATTCT AACGATAATA ATGAAAGTAA AAAAGATGAC	4620
	GCAGACAATG GTAAGAAACA AGAGATTCAA GTTGCAGCGG CAGCAAGTTT AACAGATGTA	4680
	ACCAAGAAAT TAGCTTCAGA ATTTAAAAAA GAGCATAAAA ATGCTGATAT TAAATTTAAC	4740
45	TATGGTGGAT CAGGGGCATT AAGAAAACAA ATTGAATCAG GCGCACCTGT TGACGTATTT	4800
	ATGTCTGCAA ATACTAAAGA TGTAGATGCA TTAAGAGACA AGAATAAAGC GCATGATACA	4860
	TATAAATATG CGAAAAATAG TCTAGTATTA ATTGGTGATA AAGATTCAAA TTACACTTCA	4920
50	GTAAAAGACT TAAAAGACAA TGATAAATTA GCATTAGGTG AAGTGAAAAC TGTACCAGCA	4980
	GGAAAATATG CGAAACAGTA TTTAGATAAC AATAACTTAT TTAAGAACT CGAAAGTAAA	5040

55

	CAAGGTTTTG TGTATAAAAC TGACTTATAT AAACAAAATA AAAAAATTGA TACTGTAAAA	5160
	GTAATTAAAG AAGTAGAACT TAAGAAGCCA ATCACATACG AAGCTGGTGC TACATCAGAT	5220
5	AGTAAATTAG CAAAAGAGTG GATGGAATTC TTAAAATCAG ATAAAGCTAA AGAAATACTA	5280
	AAAGAATACC ACTTTGCAGC ATAAGGAGTT GTAATCCATG CCTGACTTAA CACCTTTTTG	5340
	GATATCAATA CGAGTTGCTG TAATCAGTAC GATTATTGTA ACGGTTTTAG GTATTTTTAT	5400
10	ATCTAAATGG TTGTATCGTC GTAAGGGTTC GTGGGTAAAA GTATTGGAAA GTTTATTGAT	5460
	ATTACCTAAT GTTTTGCCGC CAACGGTATT AGGTTTTTATT CTATTAATCA TCTTCTCGCC	5520
	AAGAGGACCA ATCGGTCAAT TCTTTGCGAA TGTACTACAT TTACCTGTAG TGTTCACCTT	5580
15	GACAGGTGCT GTGATAGCAT CTGTCAATTGT TAGTTTTCCA CTAATGTATC AACATACTGT	5640
	GCAAGGCTTC AGAGGTATAG ACACGAAAAT GATTAATACA GCTAGAACGA TGGGAGCAAG	5700
20	TGAAACGAAA ATTTTCCTCA AATTAATTTT ACCATTAGCT AAACGCTCTA TTTTAGCAGG	5760
	TATAATGATG AGTTTTGCTC GTGCATTAGG TGAGTTTGGT GCTACATTAA TGGTTGCAGG	5820
	ATATATTCCA AATAAACGA ATACACTACC TTTAGAAATA TACTTCTTAG TGGAAACAAG	5880
25	TAGAGAAAAT GAAGCGTGGT TATGGGTATT AGTGCTAGTC GCATTCTCTA TTGTGGTTAT	5940
	ATCTACAATT AATTTATTGA ATAAAGATAA ATATAAGGAG GTCGACTAGA TGCTTAAAT	6000
	CAATGTGAAA TATCAATTAA AGAACCTTT AATTTCGCATC AATATAGATG ATACTGAACC	6060
30	AAAAATTTAT GCAGTTCGTG GTCCATCTGG CATTGGTAAA ACTACTGTTT TAAATATGAT	6120
	TGCCGGATTA CGTAAAGCAG ATGAAGCTAT TATCGAAGTG AATGGGCAAT TACTTACTGA	6180
	TACGGCAAAA AACGTGAATG TTAAAATTCA ACAACGACGT ATTGGATATC TGTTTCAAGA	6240
35	CTACCAATTG TTTCTAATA TGACGGTCTA TAAAAATATT ACTTTTATGG CTGAACCATC	6300
	TGAAACATC GATCAATTAA TTCAAACCTT AACATTGAT CATTGATGA AACAATATCC	6360
40	TATGACATTG TCAGGTGGAG AGGCACAACG TGTAGCACTT GCACGTGCAC TTAGCACrAA	6420
	ACCAGATTTA ATTTTATTAG ATGAACCTTT TTCTAGTTTG GATGATACTA CAAAAGATGA	6480
	GAGTATTACA TTAGTTAAAC GTATTTTCAA CGAATGGCAA ATACCAATCA TATTTGTGAC	6540
45	ACATTCAAAC TATGAAGCAG AACAAATGGC TCATGAAATT ATTACAATTG GGTAATCATT	6600
	TATTTGCCAT TAAAGAGTTT AGAACGTATT TAAAATTGTA GAAGTGAATG CTTCTATCAG	6660
	CATTTTAATG ATGTTTTAAA CTCTTTTSTA GGGGCAGTTT TTTTGAGAGA CATTGACGCG	6720
50	CGTCATATAA TGAAAGTAAT GATAAAAAGA AAGGATAACT TAATGTGAGT CAAGAACGTT	6780
	ATTCAAGGCA AATTTTATTT AAACAAATAG GTGAAATAGG TCAAAGCAAA ATAAATCAAA	6840

55

	GAGCAGGCAT TGCCAAACTA ATCATTGTTG ATAGAGATTA TATTGAATTT AGTAATTTAC	6960
	AAAGACAAAC ATTGTTTACT GAAGAAGATG CTTTGAAAAT GATGCCTAAG GTGGTTGCAG	7020
5	CTAAAAAGCA TTTGCTAGCG TTACGTAGTG ATGTTGATAT TGATGATTAT ATTGCCCATG	7080
	TGGATTATTA TTTTTTGGAA ACACATGGAC AGGACGTTGA CGTTATTATT GATGCAACCG	7140
	ATAACTTTGA AACACGACAA CTGATTAATG ATTTTGCATA TAAATATCGT ATACCTTGGA	7200
10	TTTATGGTGG TGTTGTACAG AGTACATATA CAGAAGCTGC ATTTATACCT GGTAAACAC	7260
	CTTGCTTTAA CTGTTTGGTA CCACAATTGC CAGCATTAAA TTTAACATGT GATACAGTAG	7320
	GGGTCAITCA ACCTGCCGTG ACGATGGCAA CAAGTTTACA ATTAAGAGAT GCGATGAAAG	7380
15	TATTAACGGA ACAACCAATT GACACAAAAA TAACTTATGG CGATATTTGG GAAGGTAGTC	7440
	ATTATTCATT TGGTTTCAGT AAAATGCAAC GTTCAGACTG TACAACCTGT GGAGATGTAC	7500
20	CAAGTTATCC GTATTTAAAC AAGAATGAAC AACGTTATGC AACATTGTGT GGTAGAGACA	7560
	CTGTACAGTA TGAAAATGCA TCAATTACAC ACGACATTCT TGTTCAATTT TTAAAACAAC	7620
	ATCAGTTAAA TTATCGCAGT AATTCGTATA TGGTTATGTT TGAATTTAAA GGACACCGCA	7680
25	TTGTTGCTTT TAAAGGTGGA AGGTTTTTAA TACATGGCAT GACACGCACA TCAGATGCCA	7740
	CACATCTAAT GAATTTATTG TTTGGATAAA AAAAGATAAG ACAAAGGAG TGTAATATTA	7800
	TGGGCGAACA TCAAAACGTT AAATTGAATC GTACAGTTAA AGCAGCCGTA CTAACGGTAT	7860
30	CAGATACTAG AGACTTTGAT ACAGATAAAG GTGGTCAATG CGTGCGCCAA CTATTACAAG	7920
	CAGATGACGT TGAAGTGAGT GACGCACATT ATACAATTGT GAAAGATGAA AAAGTAGCCA	7980
	TCACGACGCA GGTGAAGAAG TGGTTAGAAG AAGATATTGA TGTCATCATT ACGACTGGTG	8040
35	GAACAGGTAT TGCACAACGT GATGTGACGA TTGAAGCAGT AAAACCACTT TTAATAAAG	8100
	AGATAGAAGG CTTTGGGGAA TTGTTTAGAT ATTTGAGTTA TGTTGAAGAT GTTGGCACGC	8160
40	GTGCATTATT GTCTCGTGCT GTAGCAGGTA CAGTTAATAA TAAATTGATA TTTTCGATTC	8220
	CAGGATCAAC AGGCGCAGTT AAATTAGCAT TAGAAAAGCT CATTAAACCA GAATTAAATC	8280
	ATCTGATTCA TGAGCTTACA AAATAATTTA TTGATTTGAT TGGCGTTGAA AATCTCCAGA	8340
45	TTTACCGCCA GACTTGCTTT CAAGGTAGGT TTCGCCAATA ATCATACCTT TATCAACTGC	8400
	TTTCGTCATG TCGTAAATGG TTAAAGCCGT TGCTGATGCA GCGGTTAAAG CTTCCATTTT	8460
	AACACCGGTT TTGCCAGTTG TAGAGACAGT TGTTTGAATG TTAAAGTAT AAAGGGGTGC	8520
50	ATTTGTTTCA TCCCAGCTGA AGTGAACATC TATGCCAGTC AATGGTAATG GATGGCACAT	8580
	CGGAATAAGT GTTGATGTAT TTTTGGCAGC CATAATACCA GCGATTTGAG CAGTGTTCAA	8640

55

	AATGCTTGAA TGAGCGACAG CAGTTCCTTT TGTAATTTGT TTGTCTGATA CATCGACCAT	8760
	TTTGGCGTGG CCTTGTTGAT TAATATGAGT AACTCAGTC ATTTTACCCC TCCTAGTGCA	8820
5	TCTAGTATAT CATGAAAAA TAAAGTTTT GGAGATGATT TTTAATGGTA GTAGAAAAA	8880
	GAAACCCAAT CCCAGTTAAA GAAGCAATTC AACGTATCGT TAATCAGCAG AGTTCAATGC	8940
10	CGGCAATTAC GGTAGCACTT GAAAAAGTC TAAATCATAT CTTAGCAGAA GATATTGTAG	9000
	CTACTTATGA TATACCAAGG TTTGATAAAT CACCTTATGA TGGTTTTGCA ATTTCGAGTG	9060
	TTGATTCACA AGGGGCAAGT GGTGAGAATC GCATTGAGTT TAAAGTGATT GATCATATTG	9120
15	GTGCAGGTT AGTTTCTGAT AAATTAGTTG GGGATCACGA AGCGGTGCGT ATTATGACTG	9180
	GAGCACAAAT ACCTAATGGC GCAGATGCTG TTGTTATGTT TGAACAAACG ATTGAACTAG	9240
	AAGATACATT TACAATTCGT AAACCATTTT CAAAAATGA AAATATATCT TAAAAAGGTG	9300
20	AAGAAACAAA GACAGGCGAT GTTGTCTTAA AAAAAGGACA AGTAATTAAT CCAGGGGCTA	9360
	TCGCGGTCCT TGCAACATAT GGCTATGCAG AGGTTAAAGT TATTAAGCAA CCGAGTGTCG	9420
	CTGTATTGTC AACAGGAAGC GAATTATTAG ATGTTAATGA TGTATTAGAA GATGGGAAAA	9480
25	TTCGTAAGTC TAATGGCCCA ATGATTCGTG CCTTAGCAGA AAAATTAGGT CTTGAAGTTG	9540
	GTATTTACAA AACACAAAAA GATGATTTAG ATAGTGGCAT CCAAGTCGTT AAAGAAGCTA	9600
	TGGAAAAACA TGATATCGTT ATTACAACGG GCGGAGTTTC TGTTGGAGAT TTTGACTATT	9660
30	TACCTGAGAT TTATAAGGCT GTAAAGGCGG AAGTGTATT TAATAAAGTA GCAATGCGTC	9720
	CTGGTAGCGT AACACCGTT GCATTTGTAG ATGGaAAGTA TTTGTTGGa TTATCTGGAA	9780
	ATCCATCAGC TTGTTTTACA GGATTTGAAC TATTTGTGAA nCCAGCTGTT AACATATGT	9840
35	GTGGCGCACT AGAAGTCTTC CCGCAAATAA TTAAAGCAAC ATTAATGGAA GATTTTACCA	9900
	AGGCAAAACC ATTACACGA TTTATACGTG CTAAAGCAAC GTTAACAAGT GCTGGAGCTA	9960
40	CTGTAGTACC TTCAGGATTC AATAAATCAG GTGCGGTTGT AGCGATTGCA CATGCTAACT	10020
	GTATGGTCAT GTTACCAGGA GGGTCACGTG GTTTTAAAGC GGGGCATACA GTAGATATTA	10080
	TATTGACTGA ATCTGACGCT GCTGAAGAGG AACTTCTTTT ATGATTTTAC AAATTGTAGG	10140
45	TTACAAAAAG TCTGGTAAGA CAACATTGAT GAGGCATATT GTCTCTTTCT TAAAGTCACA	10200
	TGGTTATACA GTTGCTACTA TTAAACATCA TGGGCATGGT AAGGAAGATA TTCAATTACA	10260
	GGATTCAGAC GTCGATCACA TGAAGCATTT TGAAGCGGG GCAGATCAA GTATTGTACA	10320
50	AGGTTTTCAA TATCAGCAA CTGTAACACG TGTAGATAAT CAAATCTTA CTCAAATTAT	10380
	TGAAAAATCT GTTACAATTG ACACCAATAT CGTATTAGTT GAAGGCTTTA AAAATGCTGA	10440

55

	GAATGTTTGT TATAGCATT A ATGTAAGGGA GCATGAAGAT TTTACAGCAT TTGAGCAATG	10560
	GTTATTAAAT AAAATTAAAA ATGATTGTGA TACACAATTA ACATAGAGGA TTGAAATGAA	10620
5	TGAAACAATT TGAAATCGTG ACAGAACCGA TACAAACAGA ACAATATCGT GAATTCAC TA	10680
	TAAATGAATA TCAAGGTGCA GTAGTTGTTT TTACCGGTCA TGTTCCGCGA TGGACTAAAG	10740
	GCGTCAAAAC GGAATATTTA GAATATGAAG CGTATATTCC AATGGCTGAA AAGAAATTGG	10800
10	CACAAATTGG AGATGAAATA AATGAAAAAT GGCCTGGAAC GATAACGAGT ATTGTTCA TA	10860
	GAATAGGGCC ATTACAAATT TCAGATATCG CTGTATTAAT TCGGTTTCT TCACCGCATC	10920
	GTAAAGATGC CTATCGAGCA AATGAATATG CAATTGAGCG TATAAAAGAA ATTGTTCCGA	10980
15	TTTGGA AAAA AGAAATTTGG GAAGATGGTT CAAAATGGCA AGGGCATCAA AAAGGGAATT	11040
	ATGAAGAAGC AAAGAGGGAG GAATAAGAGA GATGAAGGTA CTTTACTTCG CAGAAATTAA	11100
	AGATATATTA CAAAAGCAC AGGAAGATAT TGTGCTTGAA CAAGCATTGA CTGTACAACA	11160
20	ATTTGAAGAT TTATTGTTTG AACGTTATCC GCAATCAAT AATAAAAAGT TTCAAGTTGC	11220
	TGTAAATGAG GAATTTGTAC AAAAATCGGA TTTCATTCAA CCTAATGATA CTGTTGCATT	11280
25	AATTCACCG GTTAGTGGAG GTTAAGGGAG CATGAAAGCA ATAATCTTG CAGGTGGTCA	11340
	TTCAAGTGC TA TTTGGTAAGC CCAAAGCTTT TCGGGAAGTG AACGGTGAGA CCTTTTATAG	11400
	TAGAGTAATT AAGACATTAG AATCAACAAA TATGTTCAAT GAAATTATTA TTAGTACAAA	11460
30	TGCGCAATTG GCAACGCAAT TTAAATATCC AAATGTTGTT ATAGATGATG AGAATCATAA	11520
	TGATAAAGGT CCATTAGCAG GAATTTATAC AATCATGAAG CAACATCCTG AAGAAGAATT	11580
	GTTTTTTGTC GTTTCTGTTG ATACACCAAT GATTACTGGT AAAGCTGTAA GCACGTTGTA	11640
35	TCAGTTTTTA GTTTCTCATC TTATTGAAAA TCATTTAGAT GTCGCAGCTT TTAAAGAAGA	11700
	TGGACGTTTT ATTCCAACAA TTGCATTTTA TAGTCCGAAT GCATTAGGCG CTATAACTAA	11760
	AGCACTACAT TCTGATAATT ACAGTTTTAA AAATGTATAT CATGAATTAT CAACGGATTA	11820
40	TTTGATGTA AGGGATGTAG ATGCGCCCTC ATATTGGTAC AAAAATATAA ATTATCAGCA	11880
	TGATTTGGAC GCTTTAATTC AAAAATTGTA AGCTGTTAGG AGGTCCACAA ATGGTAGAAC	11940
45	AAATAAAAGA TAAACTAGGA CGTCCCATCC GTGACTTACG GTTATCTGTG ACAGATCGGT	12000
	GTAAC TTTAG GTGTGATTAT TGCATGCCTA AAGAGGTATT TGGAGATGAT TTCGTATTTT	12060
	TACCTAAAA TGAAC TTTTA ACGTTTGATG AAATGGCTAG AATCGCTAAG GTATATGCAG	12120
50	AATTAGGTGT AAAAAAATA CGCATTACAG GTGGAGAACC ATTGATGCGA CGGGATTTAG	12180
	ATGTACTTAT AGCTAAATTA AATCAAATCG ATGGTATTGA AGATATTGGT TTGACTACAA	12240

55

	ATGTCAGTTT	GGATGCTATT	GATGATACGC	TATTTCAATC	AATCAATAAT	CGTAATATTA	12360
	AAGCGACTAC	GATTTTAGAA	CAAATTGATT	ACGCGACGTC	TATTGGTTTG	AATGTAAAAG	12420
5	TAAATGTTGT	TATACAAAAA	GGTATTAACG	ATGATCAAAT	CATACCAATG	CTTGAATATT	12480
	TTAAAGATAA	ACATATAGAG	ATTCGATTTA	TAGAATTTAT	GGATGTTGGT	AATGATAATG	12540
	GATGGGATTT	CAGTAAAGTT	GTAAGTAAAG	ATGAAATGCT	TACAATGATA	GAGCAGCACT	12600
10	TTGAAATCGA	TCCTGTAGAA	CCAAAATATT	TTGGGGAAGT	AGCAAAATAT	TATCGCCATA	12660
	AGGATAATGG	TGTTCAATTT	GGTTTGATTA	CAAGTGTTC	ACAATCATT	TGTTCTACAT	12720
	GTACACGCGC	AAGGCTGTCA	TCAGATGGGA	AGTTTTACGG	ATGTTTATTT	GCAACTGTCTG	12780
15	ATGGATTTAA	CGTTAAAGCG	TTTATTCGTT	CTGGCGTGAC	CGACGAAGAA	TTAAAAGAAC	12840
	AATTTAAAGC	TTTATGGCAA	ATAAGAGATG	ATCGATATTC	AGATGAGAGA	ACTGCTCAAA	12900
20	CAGTTGCCAA	TCGTCAACGT	AAAAAGATAA	ACATGAATTA	TATTGGTGGT	TAATGTGTAG	12960
	GGACCACTAC	ATATTAAATC	ATTAGAGATG	TTTTAATATT	TCTGTCTTAC	TCCCTAAAA	13020
	ACAATATTAT	TTATTAAAGT	AAAAACGGTC	ATATCTATGC	CAGATTTAAT	AGAAATGATC	13080
25	GTTTTTAAAG	TTTTTACAAG	TTGGCGGGGC	CCCAACACAG	AAGCTGACAG	AAAGTCAGCT	13140
	TACAATAATG	TGCAAGTTGG	CGGGCCCCCA	ACATAGAGAA	TTTCAAAAAG	AAATTCTACA	13200
	GACAATGCAA	GTTGGGGAAC	GGGGCCCCAA	CACAGAAGGT	GACGAAAAGT	CAGCATACAA	13260
30	TAATGTGCAA	GTTGGCGGGG	CCCCAACATA	GAGAATTTCA	AAAGAAATTC	TACAGACAA	13320
	GCAAGTTGGG	GATCAACGAA	ATAAATTTTA	TGAGAATATC	ATTTCTATCC	CACTCTTAAG	13380
	AATCACTACA	TAATAAATCT	TTAGTGGTTC	TTTAACATTG	ATGTCACACT	CCATGCCATT	13440
35	GAGTTGTAAT	ATATCTTTTT	TAGGTATAAA	TGTTGTCGAA	TAAACAACAA	GTTGTCCAAA	13500
	AGATAATAAT	CTAAACAAGA	TATAGCCAGC	AATTTAATAT	TTGTAATAGA	TAAAATGCTA	13560
	AGTTTGATAT	ATAATAAATT	TAAGTAATTG	TATAATAATA	TGAATTACAA	ACATCTAAGA	13620
40	AGAAACATAG	GAGGCATCAT	ATTATGAGTA	ATAAAGTTCA	ACGTTTTATA	GAAGCAGAAA	13680
	GGGAGTTAAG	TCAGTTAAAG	CACTGGTTAA	AAACAACACA	TAAGATTTC	ATTGAAGAAT	13740
45	TTGTAGTCCT	TTTTAAAGTG	TATGAAGCTG	AAAAGATTAG	CGGTAAAGAA	TTGAGGGATm	13800
	CATTACATTT	TGAAATGCTA	TGGGATACAA	GTAAAATCGA	TGTGATTATC	CGTAAaTCT	13860
	ATAAAAAAGA	GCTTATTTCT	AAATTGCGTT	CTGAAACGGA	TGAAAGACAA	GTATTCTATT	13920
50	TCTATAGTAC	TTCTCAAAAG	AAATTGTTAG	ATAAATTAC	TAAAGAAATA	GAAGTGTTAA	13980
	GCGTTACAAA	CTAAAACTT	aAAAagcaTG	CCAATCTCTA	TTCATCATAA	TTGCGTCTTG	14040

55

GTTCATGGCA TTTCTAGTTA CATGACGTCC ATGAATTAAG AAGTAAACAA GCATAGTAAT 14160
 GATTGCTAAA GCGGCCATAA AGCCGAAGAT TTCACTATAT GAAAACATAT GAGTAAATAA 14220
 5 CCCAAGGAAT GATGGACCGA AGCCGACACC TGCATCTAGA CCAACGTAAA AAGTAGATGT 14280
 CGCGATACCA TATTTAATCG GGGGTGAGAC TTTTATCGCA ATAGATTGCA TTGCAGATGA 14340
 TAAATTTCCA TACCCTAAAC CTAGGCAAGC ACCAGCAAGT AATATTAACC AGCTTTGATA 14400
 10 GCTTGAAATT AAGCATACAA ATGAAAGGAA AAGCATGATA AATGCTGGGT AGACAATAAT 14460
 ATTTTCATTT TTATCATCCA TCAATCTACC AGCAATAGGT CTAGTAATTA ACGATGCTAT 14520
 AGCATAGCAA ATAAAGAAAT AGCTTGCTGC AGTGACTAGG TGTCGCTCTA AAGCAAATGC 14580
 15 TTGTAAATAA GTTAGGATGG ACGCATAGGT AACGCCAATT AAAAGCATAA TTACAGCAAC 14640
 AGGAATGGCC TCTTTTGCAA TAAATTGATG AATACTAAAT CTTGGTTTAT CAATGACATT 14700
 AGTTTCAGTT TTGTTATTG TTA CTCTCGAA ATCAACTTTT ATAAATAATG AGATAATGAG 14760
 20 TCCGAGTATG CCTAATATGA CACAAATAAT AAACAGTAAG TCAATTGCGT ATTTTGTAAT 14820
 AAGTAACATG CCTAGAAATG GGCCAATCGC TGTACCTAAT ACTAACTTA AGGAAAATAA 14880
 ACTGATGCCT TCACTTTTTC TATTAACAGG GGTAACGTAT GCCGCAATAG TACCTGTTGC 14940
 25 AGTTGTCACA ACTGCAGTTG CGATACCGTT TATGAGACGT ACAAAGATTA AAAAAGCTAA 15000
 AGATCCATCA ATAAATAAAA GTAATTGCGT GATAATTAAA GCAATTAAAC CAATAAATAA 15060
 30 TAATCGTTTA GGTCCrATTT sATTTACAAA TTTACCTGTA GCAAATCGA 15109

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9072 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GAGAGTCAAT GGCAAGAAGA ATATAAATAT TTGAGAGCGT TAATCTTTAA TGAAACAGAA 60
 TTAGAGGAAG CGTATAAATG GATGCATCCT TGTTACACGT TGAATAATAA AAATGTAGTA 120
 45 CTTATCCATG GCTTCAAAAA TTATGTTGCA CTATTATTTT ATAAAGGTGC CATTTTGGAG 180
 GATAAATATC ATACACTCAT TCAACAGACT GAAAAGGTGC AAGCAGCTCG TCAGTTACGA 240
 TTTGAAAATT TAACAGAGAT TCAAGCACGT ACCGAAGAAA TTAAATATTA TCTAGCCGAA 300
 50 GCAATTAAAG CTGAAAAAGC TGGTAAAAAA GTTGAAATGA AGAAAACAGA GGAATATGTT 360

	AAATTAACGC CAGGCAGACA ACATCAATAT ATATATCATA TTGGACAAGC TAAACGCagT	480
	GgAACAAAGAC AAAAGCGTGT TGAAAAGTAT ATTAACCAAA TACTAGAAGG TAAAGGGATG	540
5	CATGATAAGT AATTAATGAG TAAAGCATAC CGGTTATACA ACAACATACA AGATGACACG	600
	AAACAACCAA TGGCTCATGC TGTGGTTGT TTTTTTAGGT GTGTCTGTCA TGGGCAACAC	660
	TTTGACGTTG GAATTCGTT ACAGGCTTGG GAGTAGAAAA TGTTAGCAA AGGCAAGGGT	720
10	GTCTACAATG AATGATGAAG ATATTAAAT ATAAGGATGA CTTTGTGAGT GGCGGATGGG	780
	CGGTTGTCCG TCTGTAACAA TGGATGCGTG TGCATTATTA CAAAAATTCG ACTTTTGTA	840
	TAATATTTCA CATTTTCGAC ACTTTTTTGC TATAAACAA CCAATTGAGC GATAATAAAT	900
15	TCGCTTTTAA AAAATATGAG TTATCTATTT AGTTGCCAAA GATAAAATAA TAATGTTTAA	960
	TAACATCATA TAGAGTATGT TAGTTTTAAA TGTCGAATAT ACGAATGTGc AAACAAAGTA	1020
	ATCGGTAGAA ATTCAACATA CATAGCGCCG TTTACTGTTA AGTATTCACA TTACAGATGA	1080
20	AAAATATAAA ATTCTACATA ATCAAGACCA TGATGTGTAC TTGTTTAACT TATGACTCTA	1140
	TTTGTTTAAc AATTGCGATA ATGGTCTTTT TATTTTATGC GTATCATTCG TCATATTTTT	1200
	TATGAGGAAG GAGAAATGAT TATGTTAAGT ATTAAGCATT TAACGAAAAT TTATTCTGGT	1260
25	AATAAAAAGG CAGTAGATGA CATCTCTTTA GATATTCAAT CTGGGGAATT TATCGCATTT	1320
	ATTGGAACCA GTGGAAGTGG CAAAACGACT GCTTTAAGAA TGATAAACCG TATGATTGAA	1380
30	GCGACAGAAG GACAAATTGA AATTGATGGT AAAGATGTTT GGAGTATGAA TCCTGTCGAA	1440
	TTGCGTAGAA ATATTGGCTA TGTTATTCAA CAAATTGGCT TAATGCCTCA TATGACGATT	1500
	AAAGAGAATA TTGTGTTGGT ACCCAAATTG TTGAAATGGA CTAAAGAGGA AAAGGATAAA	1560
35	CGTGCAAAGG AATTAATTAA ACTTGTGGAT TTACCGGAGT CATTTTTAGA GCGTTATCCA	1620
	GCAGAACTAT CAGGTGGGCA ACAACAACGT ATCGGTGTTG TAAGAGCACT TGCGGCCGAA	1680
	CAAGATATTA TTTTAATGGA TGAACCTTTT GGTGCATTGG ATCCTATTAC GAGAGATACG	1740
40	TTACAAGATT TAGTTAAAC GTTACAACGA AAATTAGGCA AGACGTTTAT CTTTGTAACA	1800
	CATGATATGG ATGAAGCGAT TAAATTAGCA GACAAAATTT GTATTATGTC AGAAGGTAAG	1860
	GTGGTGCAAT TTGATACGCC AGACAATATT TTAAGACATC CCGCAAATGA TTTTGTACGT	1920
45	GATTTTATAG GACAAAATAG ACTGATTCAA GACCGTCCCA ATGACAAGAC TGTAGAAGGT	1980
	GTAATGATTA AACCAATCAC GATACAAGCA GAAGCAACAC TGAATGACGC CGTTCATATT	2040
50	ATGAGACAAA AACGTGTTGA TACTATTTTT GTAGTAGATA GTAATAACCA TTTACTAGGT	2100
	TTCTTAGACA TTGAAGATAT AAATCAGGGT ATACGTGGAC ACAAAGTTT ACGAGACACC	2160

55

	ATTTTAAAAA GAAACGTTAG GAATGTACCT GTCGTAGATG ATCAACAGCG TTTAGTAGGA	2280
	CTGATTACGC GTGCCAATGT TGTGTGATATT GTATATGACA CGATTGGGG CGATAGTGAG	2340
5	GATACAGTGC AAACAGAACA TGTGGGGGAA GACAcTGCCT CCTCAAAAGT GCATGAGCAA	2400
	CACACTACTA ATGTCAAAGT ACGTGACATA GGAGATGATA AATCATGATT GAGTTCCTAC	2460
	ATGAACATGG TGGACAGTTG ATGTCGAAAA CACTGGAACA TTTCTATATT TCTATAGTGG	2520
10	CATTATTACT TGCCATCATT GTTGCAGTAC CTATAGGCAT TTTATTATCA AAAACAAAGC	2580
	GAAGTCCAA TATTGTATTA ACTGTGGCAG GTGTCTTACA AACTATTCCA AACTAGCTG	2640
	TACTTGCTAT TATGATACCG ATTTTGGTG TTGGTAAAAC GCCTGCAATT GTAGCGCTAT	2700
15	TTATTIATGT ATTATTACCT ATTTTAAATA ACACGGTACT CGGTGTTCAA AATATTGATA	2760
	GCAACATTAA AGAAGCTGGA AAAAGTATGG GAATGACACA ATTTCAATTG ATGAAGGATG	2820
20	TTGAATTGCC GTTAGCATTG CCGCTTATCA TTGGTGGCAT TCGTTTGTCA TCTGTGTATG	2880
	TAATTAGTTG GGCTACACTT GCAAGTTATG TAGGTGCGGG TGGATTAGGT GATTTTATTT	2940
	TCAATGGTTT AAATTTATAT GATCCACTGA TGATTGTAAC TGCAACGGTA CTCGTTACTG	3000
25	CACTAGCATT AGGTGTTGAT GCCTTATTAG CTTTAGTTGA AAAATGGGTA GTTCCCAAAG	3060
	GCTTAAAAGT ATCTGGATAA TTAGGAGGCT AAGATAATGA AGAAAATTAA ATATATACTT	3120
	GTCGTGTTTG TCTTATCGCT TACCGTATTA TCTGGATGTA GTTTGCCCGG ACTAGGTAGT	3180
30	AAGAGCACGA AAAATGATGT CAAAATTACA GCATTATCAA CAAGCGAATC GCAAATTATT	3240
	TCACATATGT TACGGTTGTT AATAGAGCAT GATACACACG GTAAGATAAA GCCAACATTA	3300
	GTAAATAATT TAGGGTCAAG TACGATTCAA CATAATGCCT TAATTAATGG GGATGCTAAT	3360
35	ATATCAGGTG TTAGATATAA TGGCACAGAT TTAACGGGAG CTTTGAAGGA AGCACCAATT	3420
	AAAAATCCTA AGAAAGCAAT GATAGCAACA CAACAAGGAT TTAAAAAGAA ATTTGATCAA	3480
	ACGTTTTTTG ATTCGTATGG TTTTGCGAAT ACGTATGCAT TCATGGTAAC GAAGGAAACC	3540
40	GCTAAAAAAT ATCATTTAGA GACAGTTTCA GATTTAGCAA AGCATAGTAA AGATTTACGT	3600
	TTAGGTATGG ATAGTTCATG GATGAATCGT AAAGGCGATG GCTATGAAGG ATTTAAAAAA	3660
45	GAGTATGGTT TTGACTTTGG TACAGTGAGA CCAATGCAAA TAGGTCTAGT CTACGACGCA	3720
	TTAAACTCAG AGAAGTTAGA CGTTGCATTA GGTATTCTA CAGATGGTCG AATTGCGGCG	3780
	TATGATTTGA AAGTACTTAA AGATGATAAA CAATTTTCC CACCTTATGC TCGAGTGCT	3840
50	GTTGCAACAA ATGAATTATT ACGGCAACAC CCAGAACTTA AAACGACGAT TAATAAGTTG	3900
	ACAGGAAAGA TTTCGACTTC AGAGATGCAA CGCTTGAATT ATGAAGCGGA TGGTAAAGGT	3960

55

	AAAGGTGGTC ATAAGTAATG GAAGGTAATT TATTACAGCA ATTATTCAAT TATTATGTTA	4080
	CGAACTTTGG TTATCTATGG GATTTATTTT TCAAACACTT ATTAATGTCT GTCTATGGTG	4140
5	TGCTGTTTGC AgCTTTAATT GGTATTCCAT TGGGAATCTT GCTTGCaAGA TACACAAAAC	4200
	TTTCTGGATT TGTAATTACA ATTGCAAATA TAATTCAAAC AGTTCCAGTC ATTGCAATGT	4260
	TAGCTATTTT AATGTTAGTC ATGGGCTTAG GTTCAGAAAC AGTAGTTTTA ACAGTGTTTT	4320
10	TATATGCGTT ACTTCCAATT ATAAAAACA CTTATACTGG TATAGCTAGT GTTGATGCGA	4380
	ATATTAAGGA TGCTGGCAAA GGTATGGGAA TGACACGCAA TCAAGTGCTA CGAATGATTG	4440
	AATTACCGTT ATCTGTTTCG GTTATTATCG GTGGCATTCTG TATTGCCTTG GTTGTTGCGA	4500
15	TAGGTGTTGT TGCCGTTGGA TCATTTATAG GAGCACCTAC GCTTGGTGAC ATTGTGATTC	4560
	GTGGTACAAA TGCGACGGAT GGCACAACGT TTATTTTAGC AGGTGCGATT CCGATTGCTA	4620
20	TCATTGCAAT CGTCATTGAT GTACTATTAA GATTTTTAGA AAAACGATTA GACCCAACAA	4680
	CACGACATCG TAAAAATCAA TCTAATCATC GGCCGCAAAG TATTAATATG TAATAGTAGA	4740
	AGATGTTTAT AATTTAGCGA TTTCGTTTCA TGATTTATAA AAAATGAGGC TACTCAAGGA	4800
25	GCTCAAATAA TCTTTGAGTA GCCTTTTTAT AGGTTGTGTT TGTATGCGTT TACACTAAAA	4860
	TAGCAATTAT TATCATGAAA GTTTTTGGAT AAAAGCGTT AATTATTGTA AAAATACTAA	4920
	AAAATGAGAT GTTTTATTTA TAATTTTCTG CAAATTATG ATATTGTTTC TTAATATATC	4980
30	ATATTAAAAA TTGTTTTTTC TTAAACATAG GAGGCTTATC TAATTCATGG ACACATCAAA	5040
	ACAATTTAGA GGTGACAACC GATTGCTTTT GGGTATCGTT TTAGGGGTTA TTACCTTTTG	5100
	GCTATTCGCG CAGTCACTTG TTAATCTTGT TGTCCCATTA CAATCAACAT ATAGTAGTGA	5160
35	CGTTGGAACG ATAAATATCG CTGTTAGCTT ATCTGCCTTA TTTGCTGGTT TGTTTATCGT	5220
	AGGTGCTGGT GATGTTGCTG ATAAATTTGG TCGCGTCAAA ATTACTTATG TAGGATTGAT	5280
	ATTAAATGTT GTAGGTTTAT TACTCATCAT CATTACACCT TTGCCAGCAT TTTTAATTAT	5340
40	AGGTAGAATA ATTCAAGGTT TGTCTGCAGC ATGTATTATG CCATCAACAC TTGCTATTAT	5400
	TAACGAATAT TATATTGGTA CAAGAAGACA ACGTGCCTTA AGCTATTGGT CTATTGGTTC	5460
	TTGGGGTGGT AGTGGTATTT GTACGTTGTT TGGTGGCTTA ATGGCTACAT ATATAGGTTG	5520
45	GCGTTCAATA TTTGTTGTTT CAATTCTATT AACATTATTA GCAATGTACT TAATCAAACA	5580
	TGCACCTGAG ACTAAAGCAG AACCAATCAA AGGTATGAAA GCAGAAGCTA AAAAGTTTGA	5640
50	CGTTATTGGT TTAGTCATTT TAGTAGTGAC GATGTTAAGT TTAAATGTAA TCATCACACA	5700
	GACGTCTCAT TTTGGTTTAG TTTCACCGTT AATCTAGGT TTAATTGTTG TGTTTATCTG	5760

55

	AATTTTAA	AATAGAGGAT	ACAGTGGTGC	AACTATTTC	AACTTCTTAT	TAAATGGTGT	5880
	AGCAGGTGGT	GCACCTATCG	TTATTAACAC	GTATTATCAA	CAACAATTAG	GATTTAATTC	5940
5	TTCGCAAACG	GGTTATATTT	CATTAAACGTA	TTTAATAACA	GTGTTGTCAA	TGATTCGTGT	6000
	AGGTGAAAAG	ATTTTATCTC	AACATGGTCC	GAAGCGCCCA	CTATTACTAG	GAAGTGGCTT	6060
	TACAGTGATT	GGGTAAATCT	TATTGTCGTT	AACATTTT	CCAGAAGTGT	GGTATATCAT	6120
10	ATCTAGTATA	GTTGGATATT	TATTGTTTGG	TACTGGTTTA	GGATTATATG	CTACACCATC	6180
	AACTGATACA	GCAGTTGCTA	GTGCGCCAGA	TGATAAGTCG	GGTGTGCTT	CAGGTGTGTA	6240
	TAAAATGGCG	TCATCATTAG	GAAATGCATT	TGGAGTAGCA	GTATCTGGTA	CGGTTTATAC	6300
15	TGTGTTAGCA	GCTAATTTAA	ATTTGAACTT	AGGTGGTTTC	ACAGGTATGA	TGTTTAATGC	6360
	CTTGCTAGCA	ATTGTTGCAT	TTTGTAGTCAT	TTTACTATTA	GTTCTTAAAA	ATCAAACGAA	6420
20	TTTGTAAAC	TGAAATGAAA	GCAAGTTATT	ATGTAGGGAT	TTTAAAGGAA	ATTTTGTGAA	6480
	AGTAAGTTTA	TCATACACAC	TTAATGTTGC	GTATTGACGT	TTAATGTTAG	GTGTGTTCTT	6540
	TTATAGACGA	TAAAAGCTGT	GTGCATATTA	AGCGAATGAT	TTTCAAATTG	ACGCTAATAT	6600
25	GCGAAAGTAG	TATTTTTTAA	ATGAACAACA	ACGATGAAGA	GGGGTTTATA	GGATGAAAAT	6660
	TGCAATTGCT	GGATCGGGTG	CATTAGGTAG	TGGCTTTGGT	GCCAAACTAT	TTCAAGCAGG	6720
	ATATGATGTC	ACACTTATTG	ACGGATATAC	ATCTCATGTT	GAAGCGGTTA	AGCAACATGG	6780
30	ATTAAATATA	ACGATTAATG	GAGAGGCATT	CGAGTTAAAC	ATTCCGATGT	ATCATTTTAA	6840
	TGATCAACCG	GACGAAAGCA	TTTACGATGT	TGTCTTTCTA	TTTCCAAAGT	CTATGCAATT	6900
	AAAAGAAGTG	ATGGAAGATA	TGAAGCCACA	TATTGATAAT	GAAACGATCG	TCGTATGTAC	6960
35	GATGAATGGT	CTGAAGCATG	AAGAAGTCAT	TGCGCAGTAT	GTTGCTCAAT	CACAAATTGT	7020
	CAGAGGTGTT	ACGACTTGGA	CGGCAGGTCT	TGAAAGCCCT	GGACACAGTC	ATTTACTTGG	7080
	TAGTGACCA	GTTGAAATAG	GTGAACTAGT	GGATGAAGGT	AAAGAAAATG	TTATAAAAGT	7140
40	TGCTGATTTA	CTTAACGAAG	CGGAATTGAA	TGGTGTCATT	AGTAAAGATT	TATACCAATC	7200
	GATTTGGA	AAGATTTGTG	TTAATGGTAC	GGCAAATGCA	TTAAGCACAG	TGTTGGAGTG	7260
	TAATATGGCA	TCGCTGAATG	AAAGTAGTTA	TGCGAAGTGT	TTGATTTATA	AATTAACGCA	7320
45	AGAAATAGTG	CATGTAGCGA	CGATTGATAA	TGTTCAATTA	AATGTTGATG	AAGTATTTGA	7380
	ATATTTAGTT	GATTTAAATG	AAaAAGTTGG	TGCGCATTAT	CCATCCATGT	ATCAAGATTT	7440
50	AATTGTTAAT	AATAGAAAAA	CTGAAATTGA	TTATATTAAT	GGCGCAGTTG	CAACATTAGG	7500
	TAAACAACGT	CaTATTGAAG	CGCCAGTCAA	TCGCTTTATT	ACTGATTTAA	TTCATACTAA	7560

55

CAATCACGTG ATATTACGGT CATTATTAAG ATTGAAATGT AATAAATAAA GAACAGCAGT 7680
 AAGGTACTTT CAAATTGAAA TGATCTTGGT GCTGTTTTTC TTGATTGATC TTCGTCATAA 7740
 5 TTCAGATTTG TCATAGGcTA CGACATACTA TTAGTATTTA CTAGACAGTT TTTACGACGA 7800
 CACTTTGAAA AATTTTGAGG CAAATCATTG GGAAGTCTCA CGTGAATTTT GTAAACTCAT 7860
 CAAGCAAGTA ATTATATTAA AAAGACAAAT AGAGAAAAGG TGTTTATAAT GAGTAAAATT 7920
 10 TTTGTAAC TG GCAACGGG CCTTATTGGC ATTAAATTAG TTCAAAGACT AAAAGAAGAG 7980
 GGGCATGAGG TTGCTGGTTT TACTACATCT GAGAATGGTC AACAAAAGCT AGCTGCTGTT 8040
 AATGTAAAAG CATATATTGG TGATATATTA AAAGCTGATA CTATTGATCA AGCGTTAGCA 8100
 15 GATTTTAAAC CAGAAATCAT TATCAATCAA ATTACGGATT TAAAAAATGT TGATATGGCA 8160
 GCAAATACGA AAGTACGTAT TGAAGTTTCT AAAAACCTAA TTGATGCGGC GAAAAAGCAT 8220
 20 GACGTTAAGA AAGTAATTGC CCAAAGTATT GCCTTTATGT ATGAACCTGG CGAAGGATTA 8280
 GCAAATGAGG AAACCTCACT TGATTTTAAAC TCAACTGGCG ATAGAAAAGT AACGGTTGAT 8340
 GGTGTGGTTG GTTTAGAAGA AGAAACGGCT CGTATGGATG AATACGTTGT TTTACGTTTT 8400
 25 GGCTGGTTAT ATGGCCCAGG TACTTGGTAC GGAAAAGATG GCATGATTTA TAATCAATTT 8460
 ATGGATGGTC AAGTGACACT TTCAGATGGC GTAACATCAT TTGTGCATCT TGATGATGCA 8520
 GTTGAAACAT CTATTCAAGC TATTCATTTT GAAAATGGTA TCTATAATGT AGCAGATGAT 8580
 30 GCACCTGTTA AAGGTTCTGA ATTTGCAGAA TGGTATAAAG AACCACTGG TGTTGAACCA 8640
 AATATTGATA TTCAACCTGC GCAACCATTG GAACGTGGCG TAAGCAATGA GAAGTTTAAA 8700
 GCGCAAGGTG GTACTCTGAT TTATCAAACCT TGGAAAGATG GCATGAATCC AATTAAATAA 8760
 35 TAATTTATCC GTTTAATATA CAAAGAATAA AGACTTGGTC GAATCGTGGA TGATATATTA 8820
 TCAAACGCAC GGCTCGAACA AGTCTTTTTT ATTATGTCTT CGTTATCTTT GTATGAAGGA 8880
 ATAACAGAAT TACAATTAAT GTACTGAATA ATGCAATTAA TGTTGTGATT AGTGCTAATT 8940
 40 TAATTTCTAT TGGTAGCCAA GTCAGTACAA AAGACCAATT ATTGCTACCG AGAATGAGAT 9000
 ATGGTAATGC ATATAATATG AGCGCTAAAG CGATACATAT ACATAATGAT AACCAACTCA 9060
 45 ATACAGCAAT CC 9072

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16826 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

	GTGGAACAGC TGTAACATA TCATTTCTTT CAACATTTAT TGGGAAAATG TTAGCTACAT	60
5	TTCTATATCC GATTAATAAT GTAGTACTTT CATATATnTC TGTAATGAA AGTGACAATA	120
	TAAAGAAGCA ATATTTGaAA ACTAATCTAA TTGCTATAGC TGCCCTATGT TTAGTCATGA	180
	TTATATGTTA TCCAATTACA ATAATTATTG TCTCTTTACT GTATAACATT GATTCAAGTT	240
10	TATATTCGAA GTTTATTATT TTAGGTAATA TAGGTGTTTT ATTCAATGCA GTGAGTATTA	300
	TGATCCAAAC TTAAATACA AAACACGCAT CAATAACATT ACAAGCGAAT TATATGACGC	360
	TTCACACGAT TACATTTATA TTCATAACTA TTTTAATGAC AATTGCGTTT GGTCTAAATG	420
15	GATTCTTTTG GACAAAGCTG TTCAGCAACA TTATTAAGTA TGTGATTTTA AATATTATAG	480
	GTTTAAAGTC TAAATTCATT AATAAAAAGG ACGTCGATTA GATGAGTGAA AAAAAGATTT	540
20	TGATTTTATG TCAGTATTTT TATCCGGAAT ATGTATCTTC TGCGACGTTA CCAACTCAAT	600
	TGGCGGAAGA TTTAATTGCG AATCACATTA ATGTCGATGT CATGTGTGGA TGGCCATATG	660
	AATATAGTAA TCATAAACAG GTTCTCTAAA CCGAGATGCA TCGTGGTATT CGCATTCGAC	720
25	GTCTCAAGTA TTCGAGGTTT AATAACAAAA GTAAGGTGG AAGGATCATC AATTTCTTTA	780
	GTTTATTTTC AAAATTCGTG ATTAATATAC CTAAAATGTT GAAATATGAT CAGATTCTTG	840
	TTTACTCTAA TCCACCAATC TTGCCATTAA TACCAGACGT TTTACACAGA CTGCTTAAGA	900
30	AAAAATATTC TTTTGTGGTG TATGATATAG CACCTGATAA TGCGATTAAG ACAGGTGCAA	960
	CTCGTCCAGG TAGCATGATT GATAAGCTGA TGC GTTACAT TAATAGACAT GTCTACAAGA	1020
	ATGCTGAAAA TGTCATTGTC CTTGGTACGG AAATGAAAAA CTACTTACTA AATCATCAAA	1080
35	TTTCTAAAAA TGCTGACAAT ATCCATGTGA TTCCTAACTG GTATGACATG CGTCAATTAC	1140
	AAGACAATCG TATCTATAAT GACACATTTA AAGCTTACCG TGAGCAATAC GACAAAATTT	1200
	TATTGTATAG CGGTAATATG GGGCAGTTAC AGGATATGGA GACACTTATC TCATTTTAA	1260
40	AATTAATAA GGATCAGTCT CAAACGTAA CAATACTTTG TGGTCATGGT AAGAAATTTG	1320
	CAGATGTCAA AACGGCAATA GaAGACCATC GTATTGAAAA TGTTAAAATG TTTGAGTTTT	1380
	TAACAGGTAC AGACTATGCT GACGTATTAA AAATGCGGA TGTATGTATT GCATCGCTGA	1440
45	TTAAAGAAGG CGTCGGTTTA GCGTGCCGA GCAAGAATTA TGGCTATCTT GCAGCTAAGA	1500
	AAGCGTTGGT ACTCATCATG GATAAGCAAT CTGATATCGT TCAACATGTT GAACAATATG	1560
50	ATGCGGGTAT CCAAATTGAT AATGGCGATG CACATGCCAT TTATAACTTC ATCAACACTC	1620
	ACTCGAGTAA GGAATTGCAC GAGATGGGTG AGCGGCACA TCAACTGTTT AAAGATAAAT	1680

55

	AAGCGATTAT TCGATGTAGT GAGTTCAATA TATGGTTTAG TAGTTTTAAG TCCGATTCTG	1800
	TTAATTACAG CATTACTAAT TAAAATGGAA TCACCTGGAC CAGCCATTTT CAAACAAAAA	1860
5	AGACCGACGA TTAATAATGA ATTGTTTAAT ATTTATAAGT TTAGATCAAT GAAAATAGAC	1920
	ACACCTAATG TTGCAACTGA TTTAATGGAT TCAACATCGT ATATAACAAA GACAGGGAAG	1980
	GTCAATCGTA AGACCTCTAT TGATGAATTG CCACAATTAT TGAATGTTTT AAAAGGAGAA	2040
10	ATGTCAATTG TAGGTCCTAG ACCAGCGCTT TATAATCAAT ACGAATTAAT CGAAAAACGT	2100
	ACAAAAGCGA ACGTGCATAC GATTAGACCA GGTGTGACAG GACTAGCTCA AGTGATGGGG	2160
	AGAGATGATA TCACTGATGA TCAAAAAGTA GCGTATGATC ATTATTACTT AACACATCAA	2220
15	TCTATGATGC TTGATATGTA TATCATATAT AAAACAATTA AAAATATCGT TACTTCAGAA	2280
	GGTGTGCATC ACTAATGAGA AAAAATATTT TAATTACAGG CGTACATGGA TATATCGGTA	2340
20	ATGCTTTAAA AGATAAGCTT ATTGAACAAG GACATCAAGT AGATCAAATT AATGTTAGGA	2400
	ATCAATTATG GAAGTCGACC TCGTTCAAAG ATTATGATGT TTTAATTCAT ACAGCAGCTT	2460
	TGGTTCACAA CAATTCACCT CAAGCAAGGC TATCTGATTA TATGCAAGTG AATATGTTGC	2520
25	TGACGAAACA ATTGGCACAA AAGGCTAAAG CTGAAGACGT TAAACAATTT ATTTTTATGA	2580
	GTACTATGGC AGTTTATGGA AAAGAAGGTC ATGTTGGTAA ATCAGATCAA GTTGATACAC	2640
	AAACACCAAT GAACCCTACG ACCAACTATG GTATTTCCAA AAAGTTCGCT GAACAAGCAT	2700
30	TACAAGAATT GATTAGTGAT TCGTTTAAAG TAGCAATTGT GAGACCACCA ATGATTTATG	2760
	GTGCACATTG CCCAGGAAAT TTCCAACGGT TAATGCAATT GTCAAAGCGA TTGCCAATCA	2820
	TTCCCAATAT TAACAATCAG CGCAGTGCAT TATATATTAA ACATCTGACA GCATTTATTG	2880
35	ATCAATTAAT ATCATTAGAA GTGACAGGTG TGTACCATCC TCAAGATAGT TTTTACTTTG	2940
	ATACATCGTC AGTAATGTAT GAAATACGTC GCCAATCACA TCGTAAAACG GTATTGATCA	3000
	ACATGCCTTC AATGCTAAAT AAGTATTTTA ATAAGTTGTC GGTCTTTAGA AAATTATTCC	3060
40	GCAATTTAAT ATACAGCAAT ACGTTATATG AAAATAATAA TGCACTTGAA ATTATTCCTG	3120
	GAAAAATGTC ACTTGTTATT GCGGACATCA TGGATGAAAC GACAACCAA GATAAGGCAT	3180
	AAGTCATCTA TTAAATAAAA TCAACATACA AATCGTTTTA TTTGGAGGTT ATAGTATGAA	3240
45	GTTAACAGTA GTTGGCTTAG GTTATATTGG TTTACCAACA TCAATTATGT TTGCAAAACA	3300
	TGGCGTCGAT GTGCTTGGTG TTGATATTAA TCAGCAAACG ATTGATAAGT TACAAAGTGG	3360
50	TCAAATTAGT ATTGAAGAAC CTGGATTACA AGAGGTTTAT GAAGAGGTAC TGTATCGGG	3420
	AAAATTGAAG GTATCTACAA CGCCAGATGC ATCTGATGTT TTTATCATTG CCGTTCCGAC	3480

55

	TAGTATTTTA TCATTTTGTAG AAAAAGGAAA TACCATTATT GTAGAGTCGA CAATTGCGCC	3600
	TAAAACGATG GATGATTTTG TAAAACCACT CATTGAAAAT TTAGGGTTTA CAATAGGTGA	3660
5	AGATATTTAT TTAGTGCAAT GTCCAGAACG TGTACTGCCA GGAAAAATTT TAGAAGAATT	3720
	AGTTCATAAC AATCGTATCA TTGGCGGTGT GACTGAAGCT TGTATTGAAG CGGGTAAACG	3780
	TGTCTATCGC ACATTCGTTC AGGGAGAAAT GATTGAAACA GATGCACGTA CTGCTGAAAT	3840
10	GAGTAAGCTA ATGGAAAACA CATATAGAGA CGTGAACATT GCTTTAGCTA ATGAATTAAC	3900
	AAAAATTTGC AATAACTTAA ATATTAATGT ATTAGATGTG ATTGAAATGG CAAACAAACA	3960
	TCCGCGTGTT AACATCCATC AGCCTGGTCC AGGTGTAGGC GGTCATTGTT TAGCTGTTGA	4020
15	TCCGTACTTT ATTATTGCTA AAGACCCTGA AAATGCAAAG TTAATTCAAA CTGGACGTGA	4080
	AATTAATAAT TCAATGCCGG CCTATGTTGT TGATACAACG AAGCAAATCA TCAAAGTGTT	4140
	GAGCGGGAAT AAAGTCACAG TATTTGGTTT AACTTATAAA GGTGATGTTG ATGATATAAG	4200
20	AGAATCACCA GCATTTGATA TTTATGAGCT ATTAATCAAA GAACCAGACA TAGAAGTATG	4260
	TGCTTATGAT CCACATGTTG AATTAGATTT TGTGGAACAT GATATGTCAC ATGCTGTCAA	4320
25	AGACGCATCG CTAGTATTGA TTTTAAGTGA CCACTCAGAA TTTAAAAATT TATCGGACAG	4380
	TCATTTTGAT AAAATGAAGC ATAAAGTGAT TTTTGATACA AAAAATGTTG TGAAATCATC	4440
	ATTTGAAGAT GTATCGTATT ATAATTATGG CAATATATTT AATTTTATCG ACAAATAAAA	4500
30	TGTGTCAAAC TAGGGCATAC ATGATTAAAG AAAGATAAGC TGTCATGTGT TTGAACCTCA	4560
	GAGAGGATAA TGTTATGAAA AAAATTATGG TTATTTTCGG TACGAGACCC GAAGCAATAA	4620
	AAATGGCACC ATTAGTAAAA GAAATTGATC ATAATGGGAA CTTTGAAGCG AACATTGTGA	4680
35	TTACAGCACA ACATAGAGAT ATGTTAGATA GTGTGTTAAG TATATTTGAT ATTCAAGCTG	4740
	ATCATGATTT AAATATTATG CAAGATCAAC AAACATTAGC AGGCCTTACG GCGAATGCAC	4800
	TTGCTAAACT TGATAGCATC ATTAATGAGG AACAACCGGA TATGATTTTA GTACATGGTG	4860
40	ATACTACAAC GACTTTTGTA GGAAGTTTGG CAGCATTTTA TCATCAAATT CCGGTCGGAC	4920
	ATGTAGAAGC TGGACTTCGA ACACATCAGA AATACTCACC ATTCCTGAA GAGTTAAATC	4980
45	GAGTCATGGT AAGTAATATT GCTGAATGA ATTTTGCGCC AACAGTAATT GCAGCTAAAA	5040
	ATTTACTTTT TGAAAACAAA GACAAAGAGC GTATCTTTAT TACTGGAAAT ACAGTTATTG	5100
	ACGCATTGTC AACACAGTT CAAATGATT TTGTTTCAAC GATTATTAAT AAACATAAAG	5160
50	GCAAGAAAGT TGTTTTACTA ACAGCGCATC GTCGTGAAAA TATTGGGGAA CCGATGCATC	5220
	AGATTTTTTAA AGCAGTAAGA GATTTGGCAG ATGAATATAA AGATGTTGTC TTCATTTATC	5280

55

	GGATTGAATT AATTGAGCCA TTAGATGCGA TTGAGTTCCA TAATTTTACA AATCAATCGT	5400
	ACCTCGTGCT GACAGATTCT GGTGGTATTC AAGAGGAGGC TCCTACATTT GGAAAACCTG	5460
5	TGTTGGTATT AAGGAATCAT ACAGAGCGTC CCGAAGGCGT TGAGGCGGGA ACATCGAGAG	5520
	TAATTGGCAC AGATTATGAC AATATTGTTC GAAATGTGAA ACAATTGATT GAGGATGATG	5580
	AAGCGTATCA ACGTATGAGT CAAGCGAATA ATCCATATGG TGATGGACAA GCATCAGAC	5640
10	GTATTTGTGA AGCAATAGAA TATTATTTTG GATTGCGCAC AGACAAGCCG GATGAATTCTG	5700
	TACCTTTACG TCACAAATAA TAAAAAAGCC CTAATCATGA AGTTGGTTTA GACAACCAGC	5760
	GGTGACTAGG GGTTTTTAAT ATATTTATTT TTGATAGTGG TAGCCAATAT CATATTTGAA	5820
15	TACTTTATTT GATAATATTG GACTTTGCTG TCCATCGTCA TCACTTTTTA AACGTACATT	5880
	TTTATGAGCT TCTTTAAATA CATCGGAATT CAACCAATTA TTAAAGCTAT CTTCAGATTC	5940
	CCAAATAGTT AAGATTTTAA CTTCGTCTGT ATCCTCGGTA TTTAATGTTT TAGTGACAAA	6000
20	CATTTGTTGG AAGCCTTCAA TAGTTTCAAT ACCTTGCTA TTGTAAAAAC GTTCAATCGT	6060
	TTCTTCCGCA CTGCCTTTTT GTAATTGTAA TCTATTTTCT GCCATAAACA TGGGCAATCA	6120
25	CTCCTCTATT TTATGATTG ATTTGGGTAA TGTTTTACA AATGTAAAGA GTACAGCGGT	6180
	TTGTATGATA ACCATTATGA TTAATCCTAC ACGGACTGCA AGAACATCCA CCATATAAAT	6240
	TGAAAAACCT ATTACAATGT ATAAGCTAAT TAAAATTTTA ATTTTCTGTT GTAGCGTGTA	6300
30	GCCTCGATGT AAATAAAAGT TTTCTACATA TTCTTTATAA ATTTTTTGAT TAATAAGCCA	6360
	ATTGTAAAAG CGATCTGAAC TTCGAGCAAA GCAAAAAACT GCTACGAGTA AAAAAGGGGT	6420
	CGTTGGCAGT AAAGGTAATA CGGCACCTGC AATACCAAGC GCTGTAAATA TTAAGCCAAT	6480
35	GACGATTAAA ATAAGTCGCA TTGAAAAAAC TCCATTCTAG TACTAATGCG CATGTAATAT	6540
	TGTTTTAGTA ATATAACTCA TGCTAAATAT AATGTGTATG ATAAGTGCAA TGA CT CAGTA	6600
	AAATGAAACG ATGTTGAATT ATCCTTGTC AATTAAACGCA TTTTAAGCGC GACTTTCATA	6660
40	ACAACCAAAC TATTTAATGA GAATTATTCT CAAGTATTAT AGTTATATTA TGTGTTTTAT	6720
	TTTTGAAAAG TGCAATATGT TTTCGAAAAT AAGATTATTT TTATGTGCAA AAACGACGCA	6780
45	AAAGTTTTAA AAATGAGACT TCTGTGAGCT GATTATTTTA TAAAATGTAA ACGCTTACTA	6840
	TATAATGTGA ATCATATCGT TTAAAAGCAT TATTAAATAT GATGCTAAGA GATTTATATT	6900
	ATAGCCAATA AACAAAGGAG AGATAATATG GCAGTAAACG TTCGAGATTA TATTGCAGAG	6960
50	AATTATGGTT TATTTATCAA TGGGGAATTT GTTAAAGGTA GCAGTGACGA AACAAATCGAA	7020
	GTGACTAATC CAGCAACTGG AGAAACACTA TCACATATTA CAAGAGCAAA AGATAAAGAT	7080

55

	TCAGAACGTG CACAAATGTT GCGTGATATT GGTGATAAAT TAATGGCACA AAAAGATAAA	7200
	ATTGCAATGA TTGAAACATT AAATAATGGT AAACCGATTG GTGAGACAAC AGCAATTGAT	7260
5	ATTCCATTTG CTGCAAGACA TTTCCATTAT TTCGCAAGTG TTATTGAAAC AGAAGAAGGT	7320
	ACAGTGAATG ATATCGATAA AGACACAATG AGTATCGTAC GACATGAGCC GATTGGCGTC	7380
	GTAGGTGCTG TTGTTGCTTG GAACTTCCCA ATGCTATTAG CTGCATGGAA GATTGCGCCA	7440
10	gCCATTGCTG CAGGTAATAC AATTGTGATT CAACCTTCGT CTTCAACACC ATTAAGTTTA	7500
	TTGGAAGTTG CTAAAATTTT CCAAGAGGTA TTACCTAAAG GTGTTGTCAA TATACTAACG	7560
	GGTAAAGGTT CAGAATCAGG TAATGCAATT TTCAATCATG ATGGTGTAGA TAAATTATCA	7620
15	TTTACGGGCT CAACTGATGT AGGTATATCAA GTTGCCGAAG CTGCAGCAAA ACATCTAGTA	7680
	CCCCTACAT TAGAGCTTGG TGGTAAAAGC GCCAATATCA TATTAGATGA TGCTAATTTA	7740
	GACCTTGCAg TTGAAGGTAT TCAGTTAGGT ATTTTATTCA ACCAAGGTGA AGTATGTAGT	7800
20	GCAGGTTCTC GATTATTAGT TCATGAAAAA ATTTATGATC AATTGGTGCC ACGTTTACAA	7860
	GAGGCATTTT CAAATATTAA AGTTGGAAAT CCACAAGATG AAGCTACACA AATGGGTAGT	7920
25	CAAACCTGGTA AGGATCAATT AGATAAAATT CAATCATATA TTGATGCAGC AAAAGAATCA	7980
	GATGCACAAA TTTTAGCAGG CGGTCATCGC TTAAGTAAA ATGGATTAGA TAAAGGGTTC	8040
	TTCTTTGAGC CGACATTAAT TGctGTGCCA GACAATCATC ACAAATTAGC ACAAGAAGAA	8100
30	ATATTTGGAC CAGTGTAAAC AGTGATTAAA GTGAAGGACG ATCAAGAAGC AATTGATATA	8160
	GCTAATGATT CTGAGTATGG TTTAGCAGGC GGTGTATTTT CTCAAAATAT CACACGTGCA	8220
	TTAAATATTG CTAAAGCTGT ACGTACAGGA CGTATTTGGA TTAACACTTA CAACCAAGTA	8280
35	CCAGAAGGCG CACCATTGCG TGGTTATAAA AAATCAGGTA TCGGTCGAGA AACTTATAAA	8340
	GGTGCGTTAA GTAACATATCA ACAAGTTAAA AATATTTATA TTGATACAAG CAATGCTTTA	8400
	AAAGGTTTGT ACTAGAATAA ATATCGTTTC TGAAGCGTGT TTGTAGGTCA GTCTAGCGGT	8460
40	AAGTCTTAAC ATTTAACGGC GTTGTTTAGA TTTTAAGCAA AACAAAATAT ATAGGAACAC	8520
	GTATCATGAT ATTAGGATAT AATGACTAAA ATAATAGCAG TAGGATGGTT TTTAATTGCA	8580
45	AATCATCTTA CTGCTGTTTT TAATTATGCT AATTTGCGAT GCGGCTATTA TAAGGACAGA	8640
	GTTGTTTTATT AATTATGGTG ATTTAGAAAT ATGAAGTTCA ATATGCAAAG TCATCGTTTG	8700
	TTTTAATATG CGGAACAATC ATTAAAGTTA TTGCGATTTT TTGAACCTAA TGAAACTAAA	8760
50	CAATAAATTT GAGATACTTT TTTGTCATTT TTATGTAAC TACACAATAA TCTCGTACAT	8820
	TATTAATAATT TTCTATATGA TAGGAATAAA GCAAAGCGCG AGTGTGCTGT AAAAGTTTTT	8880

55

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55

GATGATGTAT AAATCATGGT TAATTACGGA AGCATTAATA TTAACCTGAG AAGCTATAAA 9000
 GAATTATTTT TAAAAGCGAC AATATTAAAT ACGACGCATT TATTTAGGAG TGGCAAACGT 9060
 ATGAATGGGA AAAAGGCGAA TACGATAAAC AGATACAAAT ATTTTCATCA TGTCAATCAT 9120
 CAAAAAATTC AACAAAGTTC TAAAAAGACG CTGTGGGCAT CACTAATCAT CACATTGTTA 9180
 TTTACAGTGA TTGAATTTGT CGGAGGTTTA GTATCTAATt CATTGGCATT ACTGTCAGAT 9240
 TCATTTTCATA TGCTTAGTGA TGTATTAGCA CTTGGTTTAT CTATGTTGGC CATTTATTTT 9300
 GCAAGTAAAA AGCCGACTGC ACGATACACA TTTGGATATT TAAGATTTGA GATATTAGCT 9360
 GCATTTTAA ATGGTTTAGC ATTAATTGTA ATTTCAATCT GGATTTTATA TGAAGCTATT 9420
 GTACGTATTA TTTATCCGCA ACCAATTGAA AGTGGCATT TGTATTATGAT TGCTAGTATT 9480
 GGTTTACTCG TCAATATTAT TTTGACTGTT ATCCTTGTA GGTCTTTAAA ACAAGAAGAC 9540
 AATATCAATA TTCAAAGTGC ATTATGGCAT TTCATGGGAG ACTTATTGAA CTCTATTGGT 9600
 GTCATCGTTG CAGTTGTATT GATTTACTTT ACAGGATGGC GCATCATCGA CCCAATCATT 9660
 AGTATTGTAA TTCACTCAT CATTTTACGT GGTGGTTATA AAATTACGCG TAATGCgTGG 9720
 tTAATTTTAA TGGAAAGTGT GCCTCAACAT TTGGATACTG ATCAAATTAT GGCAGATATT 9780
 AAAACATAG ATGGCATATT AGATGTACAT GAATTTCAATT TGTGGAGTAT TACAACAGAG 9840
 CATTATTCAT TAAGTGCCCA TGTGTGTGTA GATAAAAAAT ATGAGGGTGA TGATTATCAA 9900
 GCGATTGATC AAGTATCATC ATTGTTGAAA GAAAAATATG GCATTGCACA TTCAACGTTG 9960
 CAAATTGAAA ACTTGCAATT GAATCCATTA GATGAGCCAT ACTTCGACAA ATTAACATAA 10020
 ATAAAAACATT GTAGCGCCTA AAACATTAAAT CTATGTCATA GGCGCACGTT TCGTTTTATA 10080
 CTTATGTTGC ATCATTTTAA TGATTTTCGT CAATTTCTTT GATGCTATCT ACATCTAACA 10140
 CGACtCTTT AGGTTTCAA ATATGAATAT GTTTTTCATC ATTTGTATGT AAAATGCGTT 10200
 CTATGATGTA CCTTTGACCG GCCATTGTTT CTACAGCAAT CTTTTTGTTT CTAGCTAAAC 10260
 TTGCTACGAC AGATTCTTTA TCCATAATGA TAGCCCCCTA TATATATGTT TATTTACTTA 10320
 TACCCTAACA TGATTTTTAT ACTCTTGAA AATATATTTT ACAGAATTTT ATCTAAATAT 10380
 TTAAAAAAT ATCTTAATAT CCTTGTAATC CGATAAGAAT TATAGTAATA TTTTTTCAAC 10440
 CATtGTTATA GGAGGTCTTA TTAATGACAT TATTTTATT AGAAGCTAAC AATCTTGATT 10500
 TTGCATCAAC GAAAGAAGAA CTAGAAGCAA AGGCAGCATC ACTATCTACG AAGACAATTC 10560
 CAACATTAAT TGAAGTACAA GCTACTGAAA ATTTAACTCA TGGTTATTTT ATTGTGGAAG 10620
 CAAATGACGA aGCAGAAGCT AAACAATTTT TAACAGAAGC AGATATTAGT ATTCAATTAG 10680

	TTGATTACCT TGTAACCTGG AACATTCCGG AAGGCATTAC GATGGATCAA TATTTAGCAC	10800
5	GTAAAAAGAA AAATCTGTGTT CATTATGAAG AAGTGCCAGA AGTTGAATTT AAACGCACAT	10860
	ATGTATGTGA AGATATGTCT AAATGTATTT GTTTATACAA CGCACCTGAT GAAGAAGCGG	10920
	TACGTCGCGC GCGCAAAGCA GTTGATACAC CGATTGATGG CATCGAAAAA CTTTAATAAG	10980
10	ACAACAAGTT GATGAGATAT ATGTATATAG GTTTGGCATG GATTTCGATT GCAGTTAATT	11040
	AGAATAGCTC AATGCTATAA ATGTAAGTAG TTGATATGAA GAACTAATG AACTAAATGC	11100
	AAGTATTGTC TAAAACAATC ATTTTATTGA AATTTAGTAG AGCTGAAATT AATATAACGT	11160
15	CGTTAATTGA ATAACGCTTA TGTTATAAGA GCACTCATAC CAAACCATAA TCATCTATAG	11220
	ATATAACAAT TCACGATATA AGGGCTGTGT TTGGCATAGC CCTTTAGATA TACACTTAAT	11280
	TCCTATTAAA ATAGTAGGGA TTAAAAGGGG GCTTGTCTATG ATTAAAATTC AACAATTACA	11340
20	ACATCACTTT GGATCACATA AAGTAATTCA TAACTTAAT TTGGACATTA GCAAGGGAGA	11400
	AATAGTCACT TTCATAGGGA AAAGTGGTTG CGGAAAGTCT ACTTTACTCA ATATTATCGG	11460
	TGGATTATT CATCCATCGT CTGGTCGTGT CATTATTGAT AACGAAATTA AACACAGCC	11520
25	ATCTCCAGAT TGTTTAATGC TATTTCAACA TCATAATTG CTGCCATGGA AAACGATTAA	11580
	TGACAACATT AGGATTGGAT TACAACAGAA AATTAGTGAT GAAGAGATTA ACGCACAGCT	11640
	TAAATTAGTT GATTTAGAAG ACAGGGGAAA GCATTTTCCC GAGCAACTGT CCGGGGGTAT	11700
30	GAAACAACGT GTGGCACTAT GTCGAGCGCA TGTGCATAAG CCTAACGTTA TATTGATGGA	11760
	TGAGCCATTA GGTGCATTAG ATGCATTTAC ACGTTATAAA CTTCAGGATC AACTAGTGCA	11820
	aCTAAACAT AAAACGCAAT CAACTATTAT TTTAGTGACG CATGACATTG ATGAAGCTAT	11880
35	TTATCTTTCC GACCGCATTG TTCTGTTAGG TGAAGGGTGC AATATTATTT CTCAATATGA	11940
	AATFACAGCA TCACATCCAC GCAGTCGTAA TGATAGCCAC CTACTTAAGA TTCGTAATGA	12000
	AATTATGGAA ACATTTGCAT TGAATCATCA TCAAGTTGAA CCTGAATATT ATTTATAAGG	12060
40	AGTGAGTGAC GATGAAAAGG TTAAGCATAA TCGTCATCAT TGGAATCTTT ATAATTACAG	12120
	GATGTGATTG GCAAAGGACG TCTAAAGAAC GGTCTAAAAA TGCCCAAAT CAGCAAGTGA	12180
45	TTAAAATTGG ATATTTGCCG ATTACACATT CAGCTAATTT GATGATGACT AAAAAATTAT	12240
	TATCACAATA CAATCATCCG AAATATAAAC TAGAATTAGT' TAAATTCAAT AATTGGCCAG	12300
	ATTTAATGGA CGCATTAAAC AGTGGTCGTA TTGATGGTGC ATCAACTTTA ATAGAGCTAG	12360
50	CGATGAAATC AAAACAGAAG GGCTCAAATA TAAAGGCTGT GGCATTGGGC CATCATGAAG	12420
	GCAATGTCAT TATGGGACAA AAAGGTATGC ACTTAAATGA ATTTAATAAT AATGGCGATG	12480

55

	GTAAACAATT AAAGATTAAA CCGGGGCATT TTAGCTATCA TGAAATGTCG CCAGCAGAAA	12600
	TGCCAGCCGC ATTGAGTGAA CACAGAATTA CAGGGTATTC TGTAGCCGAA CCATTCCGGT	12660
5	CACTGGGTGA AAAGTTAGGC AAAGGTAAGA CTTTGAAACA TGGTGATGAC GTTATACCTG	12720
	ATGCGTATTG CTGTGTGCTA GTACTGAGAG GGAATTGCT TGATCAACAC AAGGATGTAG	12780
	CGCAAGCATT TGTACAAGAT TATAAAAAGT CTGGCTTTAA AATGAATGAT CGCAAGCAAA	12840
10	GTGTAGACAT TATGACGCAT CATTTTAAAC AAAGTCGTGA CGTTTTAACA CAGTCAGCGG	12900
	CATGGACATC CTATGGTGTG TTAACAATTA AGCCATCCGG CTATCAAGAA ATTACGACAT	12960
	TGGTAAACA ACATCATTTG TTTAATCCAC CTGCATATGA TGACTTTGTT GAACCGTCAT	13020
15	TGTATAAGGA GGCATCGCGT TCATGACACG TCCCACAAAT AACAAATTTA TATTACCTAT	13080
	TATCACATTT ATTATTTTCT TAGGCATTG GGAATGGTC ATTATTATTG GGCATTACCA	13140
	ACCTGTATTG TTACCGGGTC CTGCTCTGT AGGAAAAAGT ATATGGTCTT TCATTGTTAC	13200
20	TGGAGAAATT TTCCAACATT TAGCAATTAG TTTATGGAGA TTTGTAGCGG GCTTTGTTGT	13260
	CGCATTGTTG GTTGCTATTC CATTGGGCTT CTTGCTTGA AGGAATCGTT GGCTATACAA	13320
	CGCTATCGAA CCGCTATTTT AATTGATTAG GCCGATATCT CCGATAGCAT GGGCACCATT	13380
25	TGTTGTTCTA TGGTTTGGTA TTGGTAGTTT GCCAGCGATT GCGATTATTT TTATCGCTGC	13440
	TTTTTTCCCA ATTGTGTTCA ATACTATTAA AGGCGTTAGA GACATTGAAC CTCAATATTT	13500
30	AAAAATAGCA GCAAATTTAA ATTTAACTGG GTGGTCATTG TATCGCAATA TATTATTTCC	13560
	CGGGGCATTT AAACAAATCA TGGCTGGGAT ACATATGGCG GTAGGAACAA GTTGATATT	13620
	TTTAGTTTCT GGTGAAATGA TTGGTGCACA ATCGGGATTA GGTTTTTTTAA TCGTTGATGC	13680
35	ACGAAATATG TTGAACCTAG AAGATGTTTT AGCAGCAATA TTCTTTATCG GATTATTTGG	13740
	TTTATTATT GATCGATTCA TTAGTTATAT TGAGCAGTTT ATACTTAGAA GATTTGGTGA	13800
	ATAAGGAGAG ATGATGATGA CTTTAGAAAC GCTTATCAAA GAACAATTAG ATCCTCATTT	13860
40	AGTAGAAGTT GATGAAGGA CGTATTATCC GAGAACATTT ATTCAGCAAT TATTTGTAGA	13920
	TGGTTATTTT GGTGAGGCGG CATTGAGAAA AAATGCTGAA GTAATCGAAG CTGTATCGCA	13980
	GTCTTGTTTG ACAACAGGAT TTTGTTTATG GTGCCAATTA GCTTTTTCAA CGTATTTAGA	14040
45	AAATGCCACG CAGCCACATT TAAATAATGA CTTACAACAG CAATTGTTAT CTGGAGAAAT	14100
	ATTAGGTGCT ACCGGATTGT CTAATCCGAT GAAGTCATTT AATGATTTAG AAAAGTTGAA	14160
50	CCTTGAACAC ACTTATGTTG ATGGACAATT GGTGTGAGT GGACGTATGC CAGCTGTAAG	14220
	TAATATTCAA GAAGACCATT ATTTTGGTGC GATTTGAAA CATGAATCAT CAGATGAATT	14280

55

	TTTAGGAGTC AACGGGTCAG CAACGTATCA AATCACATTG AATCAAGTCG TAGTGCCACA	14400
	ATCACAAATT ATCACGCATG ATGCGAAGCA GTTTGCGGCA ACTATTCGCC CGCAATTTAT	14460
5	TGCTTACCAA ATTCCAATAG GATTAGGCTC AATTAAAAGT TCTTTAGAGT TAATTGATGC	14520
	ATTTTCAAAT GTGCAAAACG GAATAAATCA ATATTTAGAG TATGATGTTG AAGCTTTTAA	14580
	AAAACGTTAT CGTCAACTTA GAGAGGAATA TTATGCAATA TTAGATGACG GTAACCTAAC	14640
10	TTCACATTTA AATGAATTAA TATCATTGAA GAAGGACATC GGCTATTTAT TGTTAGATGT	14700
	AAATCAAGCT TCTGTTGTCA ATGGTGGTTC TAGAGCGTAC ACACCATATT CGCCACAAGT	14760
	TCGCAAGTTA AAAGAAGGAT TCTTCTTCGC AGCATTGACA CCGACATTAA GACATTTAGG	14820
15	TAAACTTGAA GCAGAGTTGA AGGGGTAAGT GTGATAAGCT GATTTTTTGT TTAGATGCGT	14880
	TTGTTGAAAC ATTTTTTAAA ATAATATAAA TCTTAGTTTA TAAACATTTT CTGTTAATTT	14940
	GTTATATCCT TTAACTAGG AAAATATACA TTTCGTAATA ATAATAATCG TTATCATTGA	15000
20	AAAAGTGTTA ATAAGGTGTA TAATGAAAAT GTGAACAAAT AATGAACTTC TTATTTTAAA	15060
	GAAGGTGAAT ACTATAGATA CGCATACTAA AGAACAACAA TTCTCGAATC TAGTAAGATC	15120
	TTATCGTAAA GAATACGTGG GTAAAGGACC CAATAGTATT CGAGTGTCGT TTAAAGATAA	15180
25	TTGGGCGATT GCACATATGA CAGGTGTTTT GAGTAAAGTT GAGAGTTTTT ACCTAAACGA	15240
	CAAACGCAAT GAATCGATGC TCCATTATAC ACGCACAGAG AAGATTAAAC AGATGTATAA	15300
30	AGAAATAGAT GTAAATGAGA TGGAAAGTCT TGTAGGCGCT AAGTTTGTA AATTATTTAC	15360
	AGATATTGAT TTGAATGATG ATGAAGTCAT TTCAATATTT GTTTTCGATA AGTCAATAGA	15420
	ATAAGTGTTG CTGGTGTAAG GTACACGGTG CTGTTTGCTA ACTTCGCTTT GAATTTAACA	15480
35	ATAATTCAAG GGGGTGGTAT GTCAAACGGT GCCGTTTTTT TGTCATATTT TTAACAACAG	15540
	CAACATGCAA CACGTACTTT AAGGAAGTCA AAATTTATCA TTTAGGAGAG ATGGATATGA	15600
	AAATCGTAGC ATTATTTCCA GAAGCAGTAG AAGGTCAAGA AAATCAATTA CTTAATACTA	15660
40	AAAAAGCATT AGGATTAAAA ACATTTTTAG AGGAAAGAGG ACATGAGTTC ATTATATTAG	15720
	CAGATAATGG TGAAGACTTA GATAAACATT TACCAGATAT GGATGTGATT ATTAGTGCGC	15780
	CATTTTATCC TGCATATATG ACTCGTGAAC GTATTGAAAA AGCACCGAAC TTGAAATTAG	15840
45	CAATTACAGC AGGTGTAGGA TCTGACCATG TAGATTTAGC GGCAGCAAGT GAACACAATA	15900
	TTGGTGTCTG TGAAGTTACA GGAAGTAATA CAGTTAGTGT GGCAGAACAT GCGGTTATGG	15960
	ATTTATTAAT ACTTCTTAGA AACTATGAAG AAGGTCATCG TCAATCAGTA GAAGGTGAAT	16020
50	GGAACCTGTC TCAAGTAGGT AATCATGCGC ATGAATTACA ACACAAAACA ATTGGTATTT	16080

55

TACAACACTA TGATCCAATC AATCAACAAG ACCATAAATT GTCTAAATTT GTAAGCTTTG 16200
 ATGAACCTTGTT TTCAACAAGT GATGCGATTA CAATTCATGC ACCATTAACA CCAGAACTG 16260
 5 ATAACCTTATT TGATAAAGAT GTTTTAAGTC GTATGAAAAA ACACAGTTAT TTAGTGAATA 16320
 CTGCACGTGG TAAAATTGTA AATCGCGATG CGTTAGTTGA AGCGTTAgCA TCCGAGCATT 16380
 TACAAGGATA TGCTGGTGAT GTTTGGTATC CaCAACcTGc ACCTGCTGAT CATCCATGGA 16440
 10 GAACAATGCC TAGAAATGCT ATGACGGTTC ACTATTCAGG TATGACTTTA GAAGCACAAA 16500
 AACGTATTGA AGATGGAGTT AAAGATATTT TAGAGCGTTT CTTCAATCAT GAACCTTTCC 16560
 AAGATAAAGA TATTATTGTT GCAAGTGGTC GTATTGCTAG TAAAAGTTAT ACAGCTAAAT 16620
 15 AGAATAAGGA TGCTGGGCTA GCGATTACG CTTTCAATTT TATATAAATG AATCATATAA 16680
 GCACTACTGC TGTGTAAAG ATGGCAGTAG TTTTTTTATG ATTACATCTA AGTATAGTCA 16740
 CGGCTATGTT AGGACAATGA TTTAACATTT ACGCACATAT GTGTTCACTT ACGCAATTAT 16800
 20 TGAaAAATnT CATTCAATGTG GnaATC 16826

(2) INFORMATION FOR SEQ ID NO: 47:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4012 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTCAATGAGA GTAGTGGGCT GATGTTTAGC GATATCGCGT AAGATTAACC ATTGGCCATA 60
 35 ATATATATTG TGTTTTTCTA AAATCGGCTC GGCTAATTTT AAATAGGGGC GATATATTGT 120
 TATAAACTA TTGAAAAATT CTTGTGATAG CATAGTGACA TCTCCTAAGA CAAAATAGTT 180
 AGCTTAGCTA mCCTTTTTTAC AACAAATAGTA ATTATAAAAC GGGAGCAATT AGAAATCAAT 240
 40 ATATAATTAT TAAGAGCAAA AATAATTATA CTTTGTATAA ATAAGCGTAA TTACATGTAA 300
 ATAGGGGGAT ACTAATGATA TTGAAATTG aTCACATCAT TCATTATATA GATCAGTTAG 360
 ATCGGTTTAG TTTTCCAGGA GATGTTATAA AATTACATTC AGGTGGGTAT CATCATAAAT 420
 45 ATGGAACATT CAATAAATTA GGTATATCA ATGAAAATTA TATTGAGCTA CTAGATGTAG 480
 AAAATAATGA AAAGTTGAAA AAGATGGCAA AAACGATAGA mGGCGGAGTC GCTTTTGCTA 540
 CTCAAATTGT TCAAGAGAAG TATGAGCAAG GCTTTAAAAA TATTTGTTTG CGTACAAATG 600
 50 ATATAGAGGC AGTTAAAAAT AAACACAAA GTGAGCAGGT TGAAGTAGTA GGGCCGATTC 660

55

	ATCAGGATGA TGATGAAATT AAGCCACCAT TTTTATTCA ATGGGAAGAA AGTGATTCCA	780
	TGCGTACTAA AAAATTGCAA AAATATTTTC AAAACAATT TTCAATTGAA ACTGTTATTG	840
5	TGAAAAGTAA AAACCGATCA CAAACAGTAT CGAATTGGTT GAAATGGTTT GATATGGACA	900
	TTGTAGAAGA GAATGACCAT TACACAGATT TGATTTTAAA AAATGATGAT ATTTATTTTA	960
	GAATTGAAGA TGGTAAAGTT TCAAAATATC ATTCGGTTAT CATAAAAGAC GCACAAGCAA	1020
10	CTTCACCATA TTCAATTTTT ATCAGAGGTG CTATTTATCG CTTTGAACCA TTAGTATAAA	1080
	TATACGTAAG TGCTATGAGC GAGAATGCCC ATATGAATAA TGACAAGCAC AATGGAAAGA	1140
	ATCGTTAATA TATTATTTAA TCGTGATGAC TTAATTAAAA TGAAAAAGAT TGATAATATA	1200
15	AATGTGAAAA AGATAAGTAT AACCCGTAAA CTAAAGTAAT TCACGGTGAG AGGTTGACTC	1260
	AATGTCATAA TGATTGCAAC GATGTTTATA ATTATAAATA GACTTAAAAT AATTGTTCTC	1320
20	ATATCAAACA CCTCATTGTT AGATTATTGA CATTATAACA GGGGTAATTG TATATGAACA	1380
	TTAATGTGGT TGCTTGAGGA AAAATTTATT CATTGAAGTC AAGTTGGTTC ATTTTAGAAA	1440
	TGAATATCGT GTTAGATGAT GAAAGTATAT TGAAGTATAG GTAAGTAGTT GAAAAGTATT	1500
25	AATGTGACGA TAACATTAAA TTTAACACGA AACATAGATA TAAAATGATT CACAATTAAA	1560
	ATGGGTAAAT TTGAACCTGC TAAACTATTA ATTGGAGCAT GGACATTTCA AAAATAAGAG	1620
	TTCAAATCTT ACACAAGCTC TGAATCGACA CTATAAGATA CAAACTGTAT AATTAAAGGT	1680
30	ATTGTTAAAT AGAAGGAGAT ATCATAAATC ATGGAAAAGA TGCATATCAC TAATCAGGAA	1740
	CATGACGCAT TTGTTAAATC CCACCCAAAT GGAGATTTAT TACAATTAAC GAAATGGGCA	1800
	GAAACAAAGA AATTAAGTGG ATGGTACGCG CGAAGAATCG CTGTAGGTCG TGACGGTGAA	1860
35	G TTCAGGGTG TTGCGCAGTT ACTTTTTTAAA AAAGTACCTA AATTACCTTA TACGCTATGT	1920
	TATATTTTCG GTGGTTTTGT TGTGATTAT AGTAATAAAG AAGCGTTAAA TGCATTGTTA	1980
	GACAGTGCAA AAGAAATTGC TAAAGCTGAG AAAGCGTATG CAATTAAAAT CGATCCTGAT	2040
40	GTTGAAGTTG ATAAAGGTAC AGATGCTTTG CAAAATTTGA AAGCGCTTGG TTTTAAACAT	2100
	AAAGGATTTA AAGAAGGTTT ATCAAAAGAC TACATCCAAC CACGTATGAC TATGATTACA	2160
	CCAATTGATA AAAATGATGA TGAGTTATTA AATAGTTTTG AACGCCGAAA TCGTTCAAAA	2220
45	GTGCGCTTGG CTTTAAAGCG AGGTACGACA GTAGAACGAT CTGATAGAGA AGGTTTAAAA	2280
	ACATTTGCTG AGTTAATGAA AATCACTGGG GAACGCGATG GCTTCTTAAC GCGTGATATT	2340
	AGTTACTTTG AAAATATTTA TGATGCGTTG CATGAAGATG GAGATGCTGA ACTATTTTTA	2400
50	GTAAAGTTGG ATCCAAAAGA AAATATAGCG AAAGTAAATC AAGAATTGAA TGAACCTCAT	2460

55

	CAAAATATGA TTAATGATGC GCAAAATAAA ATTGCTAAAA ATGAAGATTT AAAACGAGAC	2580
	CTAGAAGCTT TAGAAAAGGA ACATCCTGAA GGTATTTATC TTTCTGGTGC ACTATTAATG	2640
5	TTTGCTGGCT CAAAATCATA TTACTTATAT GGTGCGTCTT CTAATGAATT TAGAGATTTT	2700
	TTACCAAATC ATCATATGCA GTATACGATG ATGAAGTATG CACGTGAACA TGGTGCAACA	2760
	ACTTACGATT TCGGTGGTAC AGATAATGAT CCAGATAAAG ACTCAGAACA TTATGGATTA	2820
10	TGGGCATTTA AAAAAGTGTG GGGAACATAC TTAAGTGAAA AGATTGGTGA ATTTGATTAT	2880
	GTATTGAATC AGCCATTGTA CCAATTAATT GAGCAAGTTA AACCGCGTTT AACAAAAGCT	2940
	AAAATTAAAA TATCTCGTAA ATTAAAACGA AAATAGATTA ACGACTGAAA TCTGAACGCT	3000
15	CATAAGACTG TCATTTGCGT TCAGATTTTT TTACACAATA TAGAATGGTT GAGTAAATA	3060
	TTTTTTGAATA TAGTGAAAGA GGGGGAAGTA CTGTGATAAA AAAGCTATTA CAATTTTCTT	3120
	TAGGGAATAA GTTTGCTATC TTTTAAATGG TTGTTTTAGT TGTCTTGGGC GGTGTATATG	3180
20	CGAGTGCTAA ATTGAAATTA GAATTACTAC CAAATGTACA AAATCCAGTT ATTTCAAGTTA	3240
	CAACAACAAT GCCGGGTGCA ACGCCACAAA GTACCCAAGA TGAAATAAGT AGTAAAATTG	3300
	ACAATCAAGT AAGATCATTG GCATATGTGA AAAATGTTAA AACGCAATCC ATACAAAATG	3360
25	CTTCAATTGT AACAGTTGAA TATGAAAATA ATACAGATAT GGATAAAGCA GAAGAACAGC	3420
	TTAAAAAGA AATCGATAAA ATTAAATTTA AAGATGAAGT TGGTCAACCA GAATTAAGAC	3480
30	GTAATTCGAT GGATGCTTTT CCGGTTTTAG CATATTCATT TTCAAATAAA GAGAATGACT	3540
	TGAAAAAAGT AACGAAAGTA CTGAATGAAC AATTAATACC AAAATTGCAA ACGGTAGATG	3600
	GTGTGCAAAA TCGCAATTA AATGGGCAGA CGAACCGTGA AATCACCTT AAATTTAAGC	3660
35	AAAATGAACT TGAAAAATAT GGGTTGACTG CTGATGATGT AGAAACTAT CTAAAAACGG	3720
	CAACAGAAC AACGCCACTT GGATTGTTCC AATTTGGTGA TAAAGATAAT CAATTGTTGT	3780
	TGATGGTCAA TATCAATCTG TTGATGCTTT TAAAAACATA AATATTCCAT TAACGTGGCA	3840
40	GGAGGACCAA GGGCATCTCA TCCCAAAGTG ACCATAAACC AAATTCAGCC ATGTCAGACG	3900
	TTATCAGGCA TCACCACAGC AAATTCAAAG CGTCAGCnCC AATATATAGT GGATGCCGCA	3960
45	nGAACTAGGG GTTTAGCGnT ATCAGTGGTG TGGCGACTCT ATTCTAAACG AT	4012

(2) INFORMATION FOR SEQ ID NO: 48:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7778 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

5	CAATATAGGT CGCCGAGTTT CAACTaCATC AACTGGTTCA GTTACATTAG ATAATGCGCT	60
	AGGTGTAGGT GGTATCCTA AAGGACGAAT TATTGAAATT TATGGTCCTG AAAGTTCTGG	120
	TAAGACAACA GTAGCGCTTC ACGCTATTGC TGAAGTACAA AGTAATGGCG GGGTGGCAGC	180
10	ATTTATCGAT GCTGAACATG CTTTAGATCC AGAATATGCT CAAGCATTAG GCGTAGATAT	240
	CGATAATTTA TATTTATCGC AACCGGATCA TGGTGAACAA GGTCTTGAAA TCGCCGAAGC	300
	ATTTGTTAGA AGTGGTGCG TGTGATATTGT AGTTGTAGAC TCAGTTGCTG CTTTAACACC	360
15	TAAAGCTGAA ATTGAAGGAG AAATGGGAGA CACTCACGTT GGTTTACAAG CTCGTTTAAAT	420
	GTCACAAGCG TTACGTAAAC TTTCAGGTGC TATTTCTAAA TCAAATACAA CTGCTATTTT	480
	CATCAACCAA ATTCGTGAAA AAGTTGGTGT TATGTTCCGT AATCCAGAGA CTACACCAGG	540
20	TGGACGTGCA TTAAAATTCT ATAGTTCAGT AAGACTAGAA GTACGTCGTG CAGAACAGCT	600
	TAAACAAGGA CAAGAAATTG TAGGTAATAG AACTAAAATT AAAGTCGTTA AAAATAAAGT	660
	GGCACCACCA TTTAGAGTAG CTGAAGTTGA TATTATGTAT GGACAAGGTA TTTCTAAAGA	720
25	GGGTGAACTT ATTGATTTAG GTGTTGAAAA CGACATCGTT GaTAAATCAG GAGCATGGTA	780
	TTCTTACAAT GGCGAACGAA TGGGTCAAGG TAAGGAAAAT GTTAAAATGT ACTTGAAAGA	840
	AAATCCACAA ATTAAAGAAG AAATTGATCG TAAATTGAGA GAAAAATTAG GTATATCTGA	900
30	TGGTGATGTT GAAGAAACAG AAGATGCACC AAAGTCATTA TTTGACGAAG AATAGTACAC	960
	AAATTTATAT CTATAGTTAA ACTTAGCAAA TATCCTTATA GGATTGATTG AAAGTGATAT	1020
	TCATCTCATA AAGCTAGAAT AATATCTAAC TTTATGGGAT ACACTACAAA TCGAGACTAT	1080
35	AAGGTTTTTT ATTTTATTIA TTATTACATT ATCAATAGTT TTATAATCGA GCTTCAAAC	1140
	TTTAGAAAAAT AGTAGAAATA GCATTCAATA TAGTGCAAAA GTGCAAATTG ATAACCTGAC	1200
	ACTTATCTCC TATAAACCGT ACAATTAATT TGTATGATTT ATATATAATT TCATAAAGTC	1260
40	ATATTGAATT TCATATAAAG AGCAAACCCT AGAAAAGGAG GTGTTTGTGT GAATTTATTA	1320
	AGCCTCCTAC TCATTTTGCT GGGGATCATT CTAGGAGTTG TTGGAGGGTA TGTGTTGCC	1380
45	CGAAATTTGT TGCTTCAAAA GCAATCACAA GCTAGACAAA CTGCCGAAGA TATTGTAAAT	1440
	CAAGCACATA AAGAAGCTGA CAATATCAAA AAAGAGAAAT TACTTGAGGC AAAAGAAGAA	1500
	AACCAATCC TAAGAGAACA AACTGAAGCA GAACTACGAG AAAGACGTAG CGAACTTCAA	1560
50	AGACAAGAAA CCCGACTTCT TCAAAAAGAA GAAAACCTAG AGCGCAAATC TGATCTATTA	1620
	GATAAAAAAG ATGAGATTTT AGAGCAAAAA GAATCAAAAA TTGAAGAAAA ACAACAACAA	1680

55

	CGCATCTCCG GTCTCACTCA AGAAGAAGCT ATTAATGAGC AACTTCAAAG AGTAGAGGAA	1800
	GAACTGTCAC AAGATATTGC AGTACTTGTT AAAGAAAAAG AAAAAGAAGC TAAAGAAAAA	1860
5	GTTGATAAAA CAGCAAAAGA ATTATTAGCT ACAGCAGTAC AAAGATTAGC AGCAGATCAC	1920
	ACAAGTGAAT CAACGGTATC AGTAGTTAAC TTACCTAATG ATGAGATGAA AGGTCGAATC	1980
	ATTGGACGAG AAGGACGAAA CATCCGCACA CTTGAACTT TAACTGGCAT TGATTTAATT	2040
10	ATTGATGACA CACCAGAAGC GGTTATATTA TCTGGTTTGT ATCCAATAAG AAGAGAAATT	2100
	GCTAGAACAG CACTTGTTAA CTTAGTATCT GATGGACGTA TTCATCCAGG TAGAATTGAA	2160
	GATATGGTCG AAAAAGCTAG AAAAGAAGTA GACGATATTA TTAGAGAAGC AGGTGAACAA	2220
15	GCTACATTTG AAGTGAACGC ACATAATATG CATCCTGACT TAGTAAAAAT TGTAGGGCGT	2280
	TTAACTATC GTACGAGTTA CGGTCAAAAT GTACTTAAAC ATTCAATTGA AGTTGCGCAT	2340
20	CTTGCTAGTA TGTTAGCTGC TGAGCTAGGC GAAGATGAGA CATTAGCGAA ACGAGCTGGA	2400
	CTTTTACATG ATGTTGGTAA AGCAATTGAT CATGAAGTAG AAGGTAGTCA TGTTGAAATC	2460
	GGTGTAGAAT TAGCGAAAAA ATATGGTGAA AATGAAACAG TTATTAATGC AATCCATTCT	2520
25	CATCATGGTG ATGTTGAACC TACATCTATT ATATCTATCC TTGTTGCTGC TGCAGATGCA	2580
	TTGTCTGCGG CTCGTCCAGG TGCAAGAAAA GAAACATTAG AGAATTATAT TCGTCGATTA	2640
	GAACGTTTAG AAACGTTATC AGAAAGTTAT GATGGTGTAG AAAAAGCATT TGCGATTCAG	2700
30	GCAGGTAGAG AAATCCGAGT GATTGTATCT CCTGAAGAAA TTGATGATTT AAAATCTTAT	2760
	CGATTGGCTA GAGATATTAA AAATCAGATT GAAGATGAAT TACAATATCC TGGTCATATC	2820
	AAGGTGACAG TTGTTGAGGA GACTAGAGCA GTAGAATATG CGAAATAATT TTTGTCTCCC	2880
35	TCACAAATTA GTGAGGGAGC TTTTTTAAGT TGTAGTCTTA ACTAGTTAG ACAGCACTTT	2940
	ATCGGTAATA ACTATATTAA ACAGTAGTTA TTTGAAAGTA AGACGGACCT TATATTAAAT	3000
	AAGAAGTTAT TGCTTTTAAAT AAAAATGTTT TAGGCTTCGT AATTACTATA TTTATATTAT	3060
40	GTAAACCTAT AAAGATGATT GGTTTTCTAT CCAATAAAAA AGAAGAGAAG ATGTAACACA	3120
	TCTTCTCTTC TGCAATATTA ATTAGGATTT ATTTCTAAGT TGAGTTATTT TAATTGTAAA	3180
	TCTGTTTTCT TTAATCTTT TATAACTTCT GCAGTATCAT AACAATTTGT TGCAATTGTT	3240
45	GAATATCTCT CTGCTAAACG ATATGCATTA ATGTAAAGCT TTAAACTTTC TTTAGCTATA	3300
	TCCTCTGCAT CTTGCAATTT TGATGGGTTA GACATAACCA CTAATCTGTC AAATTTTCT	3360
50	GGATCAATAT TAATAGACAT GTATTTATTT ACAACTCCTA TTTATTTTGA TGTCTTAATA	3420
	CTAACATATT GAAGTTTTCA GACAAAGTAA TGTCTCTCTA TAATTGAAGA AAAATAATTC	3480

55

	GGATGAACAA AACATGAGAA TAATGTTTAT AGGGGATATC GTAGGTAAAA TTGGACGAGA	3600
	CGCAATTGAA ACGTACATAC CTCAACTGAA GCAAAAAGTAT AAACCAACAG TTACAATTGT	3660
5	AAATGCTGAA AATGCAGCAC ATGGTAAAGG TTTGACTGAA AAAATATATA AACCAATTACT	3720
	AAGAAATGGT GTAGATTTCA TGAATATGGG TAATCACACA TATGGTCAAC GTGAAATTTA	3780
	TGATTTTATA GATGAAGCAA AACGACTAGT AAGACCAGCG AATTTTCCGG ATGAAGCGCC	3840
10	GGGAATTGGT ATGAGATTTA TACAAATTAA TGATATTAAA CTTGCAGTTA TTAATCTGCA	3900
	AGGAAGAGCG TTTATGCCAG ATATTGATGA TCCTTTTAAA AAGGCAGATC AATTAGTCAA	3960
	GGAAGCACAA GAACAAACTC CGTTTATATT TGTTGATTTT CATGCAGAAA CAACTTCTGA	4020
15	AAAGTATGCA ATGGGATGGC ATTTAGATGG TAGAaTAGCG CTGTTGTTGG AACGCATACA	4080
	CACATTCAAA CAGCAGATGA ACGTATTTTA CCAAAGGGA CAGGGTATAT AACGGATGTT	4140
	GGTATGACAG GTTTTTATGA TGGCATTTTA GGAATAAATA AAACAGAGGT AATTGAGCGT	4200
20	TTTATCACTA GTTTGCCACA AAGACATGTT GTTCCAAATG AAGGTAGAAG TGTATTATCT	4260
	GGTGTGTGTA TTGATTTAGA CAAAGAAGGT AAAACAAAGC ACATCGAACG TATATTGATA	4320
25	AATGATGACC ATCCATTTTC AACATTTTAA AATTACGTAA GTAAACATTG GAATTGGACC	4380
	CTATCGTCCA TTAGTATGAA TTTAATATAG TACCACTGTT TACATAGTAA ATCGGTGGTT	4440
	CTTTTTGTGA TCATTTAATA TGAAATATAT CCATAGGAGG CATATAACTA TGAAACCACA	4500
30	ATTATCGTGG AAAGTTGGCG GTCAACAAGG CGAAGGTATT GAATCAACTG GGGAAATCTT	4560
	CGCTACGGCT ATGAATAGAA AAGGATATTA TTTATATGGA TATAGACATT TTTCAAGTCG	4620
	TATCAAAGGT GGACATACGA ATAATAAAAT TAGAGTTTCT ACGACGCCTG TTCATGCAAT	4680
35	TAGTGATGAT TTAGATATTT TGATTGCATT TGACCAAGAA ACAATTGATG TTAACCATCA	4740
	TGAAATGAGA GAAGACAGTA TTATTTTArC TGATGCCAAG GCTAAACCTG TGAAaCCAGA	4800
	AGGATGTCAT GCACAGCTTA TTGAATTACC TTTTACAGCA ACCGCTAAAG AATTAGGTAC	4860
40	AGCATTAAATG AAAAACATGG TTGCAATAGG TGCTACTAGC GCATTGATGA ATTTGAATAC	4920
	AAATACATTT GAAGAACTTA TTACTAATAT GTTTTCTAAA AAAGGTGACA AGGTAGTTGA	4980
	AGTCAATATC CAAGCATTAA ACGAAGGTTA TCAATTAATG CAATCTCGCT TACCTGAAAT	5040
45	CTACGGGGAC TTTGAATTAG AGTCAACAGA TGCACTACCA CATCTATATA TGATTGGTAA	5100
	CGATGCCATT GGATTAGGTG CAATTGCTGC AGGTTCCAAA TTTATGGCGG CATATCCTAT	5160
50	TACACCTGCG TCTGAAGTTA TGGAATATAT GATTGCCAAT ATATCTAAAG TAAACGGAGC	5220
	GGTTATTCAA ACAGAAGATG AAATTGCTGC TGTAACATATG GCTATTGGTG CAAATTATGG	5280

55

	TGGATTATCT	GGTATGACTG	AAACGCCATT	AGTCATTATT	AATACCCAAC	GAGGTGGACC	5400
5	TTCTACTGGA	TTACCTACGA	AACAAGAACA	GTCAGATTTA	ATGCAAATGA	TTTATGGTAC	5460
	ACATGGTGAT	ATTCCAAAAA	TTGTTGTAGC	ACCAACAGAT	GCAGAAGATG	CATTTTATTT	5520
	AACTATGGAA	GCATTTAATT	TAGCAGAACA	ATATCAATGC	CCTGTTATAG	TTCTAAGTGA	5580
10	TTTGCAATTA	TCTTTAGGTA	AACAACTGT	TGAAAAATTA	GATTATAATC	GTATTGAAAT	5640
	TAAACGTGGT	GAAATCATT	AATCTGATAT	TGAACGTGAA	GAAGATGATA	AAGGTTATTT	5700
	CAAGCGTTAT	GCGTtAACAT	CCGATGGTGT	TTCTCCTAGA	CCTATCCCCG	GTGTTAAAGG	5760
15	AGGTATTCAT	CATATAACTG	GTGTGGAAcCa	CAATGAAGAA	GGTAAACCTA	GTGAATCTGC	5820
	GTCAAATAGA	CAACAACAAA	TGGAAAAACG	AATGCGTAAA	ATTGAGCAGT	TACTAATTGA	5880
	ATCGCCAGTA	GAAGCTAACT	TACAACATGA	GGATGCAGAT	ATTCTTTATA	TCGGTTTTAT	5940
20	TTCTACAAAA	GGTGCAATTC	AAGAAGGTAG	TAACCGTTTG	AATCAACAAG	GCATAAAAGT	6000
	TAACACTATA	CAAATTAGAC	AATTGCATCC	ATTCCCAACA	AGCGTTATTC	AAGATGCAGT	6060
25	TAATAAAGCG	AAGAAAGTCG	TTGTAGTGGA	GCACAATTAT	CAAGGACAAT	TGGCTAGTAT	6120
	TATAAAAAATG	AATGTCAATA	TTCATGATAA	GATTGAAAAT	TATACAAAGT	ATGATGGGAC	6180
	ACCTTTCCTA	CCACATGAAA	TCGAAGAAAA	AGGCAAAATA	ATTGCTACTG	AAATAAAGGA	6240
30	GATGGTATAG	ATGGCGACAT	TTAAAGATTT	TAGAAATAAT	GTTAAGCCTA	ACTGGTGCCC	6300
	CGGATGTGGC	GATTTCTCAG	TACAAGCTGC	AATTCAAAAA	GCAGCCGCAA	ATATAGGGTT	6360
	AGAACCTGAA	GAAGTAGCTA	TCATCACCGG	TATAGGATGT	TCTGGCCGTC	TTTCAGGATA	6420
35	TATTAATTCT	TATGGCGTTC	ATTCTATTCA	CGGACGTGCA	TTACCTTTAG	CTCAAGGTGT	6480
	AAAAATGGCG	AATAAAGATT	TAAGTGTAT	TGCATCGGGA	GGAGATGGTG	ATGGTTATGC	6540
40	TATAGGTATG	GGGCATACAA	TCCATGCTTT	AAGAAGAAAT	ATGAACATGA	CGTATATAGT	6600
	CATGGATAAT	CAAATTTATG	GTTTGACAAA	GGGACAAACA	TCGCCGTCAT	CAGCAGTAGG	6660
	ATTTGTTACT	AAAACAACGC	CAAAAGGTAA	TATAGAAAAA	AATGTTGCGC	CTTTAGAATT	6720
45	AGTATTATCA	TCTGGTGCCA	CATTTGTAGC	CCAAGGTTTT	TCAAGCGATA	TTAAAGGATT	6780
	AACAAAACATA	ATTGAAGATG	cAATTAATCA	TGATGGATTT	TCATTGTTA	ATGTCTTTTC	6840
	ACCATGTGTG	ACTTATAATA	AAATTAACAC	ATACGATTGG	TTTaAAGAAC	ATTTAACAAG	6900
50	TGTTGATGAc	ATTGAAAATT	ATGATTCTAC	AGATAAACAA	TTAGCGACTA	AAACTGTTAT	6960
	TGAACATGAA	TCTTTAGTAA	CTGGTATTGT	TTATCaAGAT	AAAGAAACAC	CATCATATGA	7020
55	ATCtCAAATT	AAAGAGTTAG	ATGATmCACC	ACTTGCTAAA	AGAGATATCa	AAATTaCTGA	7080

TGTATTTATA ACAGATCCAT TTATGCTACT CAGTTTTTTA CTATTACAAA AAATAAAGGA 7200
 GTTTTTAAAA ATGAAAGACA CATTAAATGAG TATACAAATA ATTCCTAAAA CACCAAACAA 7260
 5 TGACAATGTT ATACCTTACG TAGACGAGGC GATTAAAAATA ATTGACGAAT CTGGTTTGCA 7320
 TTTTAGAGTA GGTCCGTTAG AAACGACAGT ACAAGGAAAT ATGAATGAAT GTTTAATTTT 7380
 10 AATACAATCA TTAAATGAAC GAATGGTGGA ACTTGAATGT CCAAGTATTA TTAGCCAAGT 7440
 TAAGTTTTAT CATGTGCCAG ATGGCATCAC TATTGAAACT TTAAGTAAA AATATGATGA 7500
 ATAACATTAA AAGTGAAGTA AACTGGATTT GAATTGGCTT GTTAGAGATG ACGTATAACT 7560
 15 TTAAGTGTTC TTGCACITTA TAGTTAAATT TAATATAATT ATTAAATGAT ACGGGCAAAT 7620
 AGAAAGGATT TTGTAAAGTG AACGAAGAAC AAAGAAAAGC AAGTTCTGTA GATGTTTTAG 7680
 CTGAGAGAGA TAAGAAAGCA GAAAAAGATT ATAGTAAATA TTTTGAACAT GTTTATCAGC 7740
 20 CGCCTAATTT AAAAGCAAGC GCAAAAAAAG AGGTnAAA 7778

(2) INFORMATION FOR SEQ ID NO: 49:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGATGAAGTT GTTACgAAAA TTGCGTACGC TGTTTCAGAA CATGTCAAAA TAGAAACAGG 60
 35 TAATCCATTTC TTTCAAACAT CACATAGTGG TTGTGCGACG GCGGATCCT GTAATTGTTC 120
 ATTATAAAAA ACATCGAGTC AGAAAAAGGT GGTATTGAA cCACTAACTA GCATCTGACT 180
 40 CGATGTTTTT ATTTATTGGG GATTGTTTGT TTGAATTGTT GTGCTAAATC TGGTCGATCT 240
 GTCACAATCG TGTGTGCACC TTTTGGTAT AAATCATTCA TCAGATTTAT ACTATTTACG 300
 CCATAATAGC CTGGAATGAT ATTCATATCA TTAAACCATT TGATAAAACG AGATGAAGTC 360
 45 AAATCAATGC CTTTAAAATG AGTAGGCATT TGAACGTTT GTGCTAATGG TTGGTAGTAC 420
 CTACCACCTA ATAAATGATA TTTTAAAAAT GCTTCTGTAA CTTCTGTGTT GCTAGACCA 480
 ATTGCGACGG ATCCTTGTGC AATTTTATTA AAACGAACGA TTTGTTCTTT ATAAAACTT 540
 50 GTCACAAGAA CGCGGTCAAA TGCTTGATTT TCTGCAATTG TATCAAACAT AATTGTGGT 600
 GCGATTGAGC CTTCATAGGA TTCAGGAGCA TCTTTTAAGT CTACGTTTAT ATACATATCA 660
 55 GGATATTGCT TCAGCAACTc ATCGAAGGTT AGTATAGCTG TGTGTGCATG ACCACGATAT 720

AATGTATGGG CACTAACTTT TCCAGAGCCG TTCGTCGTTT TATCAACAGT TGCATCATGA 840
 AAAACGATAA GCTGTTGATC TTTTGTGAGT CTCACATCTG TTTCAAAGCC ATCAACGCCT 900
 5 AATTGTTTAG CATAGTCAAA TGCAAGTTGC GTTTGCTCTG GTCTTAAAGC CATACCACCG 960
 CGATGCGCAA ATATATATGG TGCATTGCCT TTGAAAAAAG CAGGGATGGT TTGCTTTTTA 1020
 10 GTAATCACTT TATTTTTATT GATCATTAAT AGACTACTTA AAAATCCAGC ACCGACTAGT 1080
 ACCGCATTTA AAATGTTTCT GTTTACnTTT TTCATAAAAA ATTCCTCC 1128

(2) INFORMATION FOR SEQ ID NO: 50:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6252 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

25 CAAGCAAACA ATCGTCGATA AAATTGCTAA AATAATAAAA GTAATTCGAA CTTTCATCAT 60
 GATCATCCTT TGTTTATAGA GTCAATATAA GTATGGAATA TGTTAGGTAT ATAGTCAAAT 120
 GCGTCAACTA ATGGGAATTT TGGCATAGAT AGAGAATTTA AGGCAATTAA AAAGGCATCA 180
 30 AACAGTAATA TGCTGCTTGA TGCCCAAATG ATGACTTTAG CTAAATTGAT TAGTCACTTT 240
 TAAAGATAAA GAATTGTCAT GAATTAAAAC TCATGTAATG ATGTGTTACA TTTCGCAATG 300
 ATGGCTTTCA GTTATTTATC GATAACATCA CTCTTGATAC CTTTAGATTT TAAGAAATCT 360
 35 TTAATTTTAT CTGTGTTGCTT TTTATTAACA TCACCGGCAT ATTTGTTGG CACGTCGACA 420
 ACATTGATTT TATTTTGCGG TTGATAGCTA AGCTTTTCAA TATCTTCATC AACATTGGCG 480
 ATTGTTACTAT TTAAAGCTTT GAAGTAATTC ATCATTAATT CAACGGGTTT CTTATATTCT 540
 40 TTAGGAATAT TGTTTTCAGT GACAAATTTT TTGAAATGCA AATCGTTTTT AACAGCTAAG 600
 TTAGATAAGT GGCTAAGTGT TTCTGCTTGT TTTTCAGTCA CTTTGTGTTG ACTGTCAATT 660
 45 TGTTTATCTA GTTTATGTTG CATAATATAT TTGTTATCAA GTATATCGCT ATTTACAGAC 720
 AAATACTTTT CTATAGCTTG CTTTCATCTCT GCATCACTAA TATCACTATT TTTCTTATCT 780
 GAGTTAAAGA TATCTTTTGT tTCTAATTTT TTAGCGCTTT TAGGTGCATG GATGCCAGTA 840
 50 CTTGTATGAT GATCTTCGTT ATCAGATTGA TCGGACGCGC AACCTGTAAG AATTAATGTC 900
 GATGCTAAAA ATGTACTTAG TAGTAATCTC TTTTTCATAA TGTAATATAA CTCCTTAGTT 960
 55 TATCTTTAAT TGAAAAAATA TGTATTCATG TTTAATAGAG TAACATTGAA TTAGTTTGGA 1020

	TCTATCAATA ATGCATCATT TTGGACGTTG TTAAGGATAG CTTTATCTAT AAATAACTGC	1140
	ATAATTGGTT GTRACTAATTT AGACGTAGGT ATCGTACGTA AAAGCATAAT AATTTTCGTT	1200
5	ACATACTTTT CTTTCTCAAT ATCATTITTTT ATATTGATTT GTTTGCGAGA GGTACATACT	1260
	TTAAGCATT TCGCACATCT CGTTGTATAT ATTAAGTTTA TCATAACATG ATTTTATGTC	1320
10	GGGATAAAAA AATAACAGCA TCTTAACAAA TGTAAGATAC TGTCAAGTAA ATGAATGAAA	1380
	CTTTAGTTTC TGaTAATATA GTCAAAGGCA TTTAATGCTG CATTTGCACC AGCGCCCAT	1440
	GAAATGATAA TTTGTTTGTT CTTCTGATCT GTGACATCGC CAGCAGCAAA TATTCCAGGA	1500
15	ACATTCGTAT TATTGTTACG ATCAATCACA ATTTACCAC GTTCGTTTAA TTCAACAGCA	1560
	TCGTTTAAAC ATGATGTGTT TGGAAGTAAA CCAATTTGAA CAAAGATACC ATCTAAGTTA	1620
	AGTAGATGTT CTTGCGCGGT GTTCATGTCT TCGTAACGTA TACCTGTAAC ATGGTCTTCT	1680
20	CCGACAACCT CAGTAGTTTT GGCATTTGTT TTGATATCAA CATTTGATAA AGAACGTAAA	1740
	CGATCTTGTA ACACGTTGTC TGCTTTTAAT TCGCTAGCGA ATTCGAATAA TGTAACATGA	1800
25	TTAACGATAC CAGCAAGGTC AATTGCTGCT TCAACCCAG AGTTACCGCC ACCGATAACT	1860
	GCTACGTCTT TATTTTCAA TAGAGGTCCTG TCACAGTGAG GGCAGAATGC AACACCTTTA	1920
	TTAATCAATT GCTCTTCACC TGGAATGTTT AGCTTACGCC AACCTGCACC AGTAGCAATA	1980
30	ATGACTGTTT TACTTTCTAA GACAGCACCG TTTTCTAACG TAACTTTAAAT TGCTTCGTCA	2040
	GTCTTTTCGA TATCTGTAGC ACGTATACCT GTCATTGCAT CAATGTCATA TTGATCAATG	2100
	TGCGCTGCTA AGTTAGAAGA AAATTCAGAA CCAGTTGTTT CTTTAACAGT AATGAAGTTC	2160
35	TCAATACCAG CAGTATCATT AACTTGGCCA CCGATACGAT CAGCAACTAT ACCAGTACGT	2220
	AAACCTTTAC GTGCTGTGTA AATCGCTGCA CTACCACTAG CAGGACCACC ACCAACGATT	2280
40	AAGACATCAT AAGGTTCTTT ATTTTCAAAC TCAGATGCAT CTGCCGTACT GCCTAGTTTC	2340
	GAAAGAATAT CTTGGATTGT CATACGACCA TTGCCAAATT CTTGCCATT TAAAAAGACA	2400
	GCAGGGACTG CCATGATGTT TTCAGATTCT TCACGGAACA CTGCACCATC AATCATAGAA	2460
45	TGCGTGATGT TAGGGTTGAT CACACTCATT AAGTTAAGTG CTTGAACGAC ATCAGGACAT	2520
	TTTTGACACG TTAAACTAAT GAATGTTTCA AAATGGAATG AACCTTCTAA TTTTTTAATT	2580
	TGGTCAATGA TTGACTGTTT TTCTTTAGGT GCACGACCAC TAACCTGTAA AATTGCTAAA	2640
50	ACAAGTGAGT TAAACTCGTG ACCTAATGGA ATACCTGCAA ATGTTACACC TGTTTCTTCG	2700
	CCAGGACGAT TGACTGAGAA ACTTGGTGTA CGTTTTAAAG ATTTTTCAGA AAGAGATAGT	2760
55	CTAGGTGACA TATCAGTAAT TTCTGTCAAC AAATCTTTAA GTTCTTTGGA TTTATCATCT	2820

	TGTTGTTTTA AATCAGCATT AAGCATGGTT GTAATGCCTC CTTAGATTTT ACCTACTAAA	2940
5	TCTAAACCAG GTTGCAATGT TTTAGCGCCT TCTTCCCATT TAGCTGGGCA TACTTCGCCA	3000
	GGGTTTTTAC GAACATATTG AGCTGCTTTG ATTTTGTGAG CTAATGTACT AGCGTCACGG	3060
	CCAATTCGGT CAGCGTTAAT TTCAGATGCT TGTACAACAC CGTCTGGGTC GATAATGAAT	3120
10	GTACCACGTT GAGCTAAACC AGTAGCTTCA TCTAATACAT CAAAATTACG AGTGATTGTT	3180
	TGTGATGGGT CACCAATCAT AGTGTAAGTG ATTTTGCTAA TTGCATCTGA ATGGTCATGC	3240
	CATGCTTTGT GTACGAAGTG AGTATCAGTT GATACTGAGA ATACATTTAC GCCTAATTTT	3300
15	TGTAATTCTT CATATTGGTT TTGTAAGTCT TCTAATTCAG TTGGACAAAC GAATGAGAAG	3360
	TCAGCAGGAT AGAAGCATAC TACGCTCCAA GAACCTTTTA AATCTTCTTG TGTAACCTCT	3420
	TTAAATTGAT CTTTTTTTGG ATCGAAARCT TGCCTGTAA ATGGTAAGAT TTCTTTGTTA	3480
20	ATTAATGACA TAAATATCTT CCTCCTAAGA ATTTAAGTAT GAATTAGAAC TATCAATTGA	3540
	TTGCGCTTAA TTATAATAAT TCTAATCTCT TAGTTAGCAT TATTACATTT TGATCCAGAA	3600
25	TAGTCAACTG GATAACTTTG TAAAGTGAAT GATTACTTTT AAAATAAAGA AAGATAATAT	3660
	AAAGTGCTTT GATAATGGAT TTTGTAGTTG ATGATTTAAA AGGTTGTGTC TATATTTAAT	3720
	ATCTTGATTT TAATGTAAAA AATGTAAAAA AAGAAGATTT GTATTCTCAA CTAAGTCAAC	3780
30	CTTATTGATA ATGGTATGAG AATATTTGTT CGAGATGGAT GAAGGTAATG AGTGAGAAAC	3840
	TGGATTTTTA AAGTATGAGA CAATATTTTA AAAAGTTCAA TTATTAACCTT ATAAGCAAAT	3900
	AATTGCTATA AAAAAGTTTG GACGTGTACA ATTGCAATAT GAAGATTTTA AATTAATTGT	3960
35	AAAGTATCGA GGAGTGGGTA ACGTGTGAGA ACATGTATAT AATCTTGTGA AAAAGCATCA	4020
	TTCTGTTAGA AAATTTAAGA ATAAACCTTT AAGTGAAGAC GTTGTTAAGA AATTGGTAGA	4080
40	AGCTGGACAA AGCGCTTCGA CGTCAAGTTT CCTGCAAGCA TACTCAATTA TTGGTATCGA	4140
	CGATGAGAAG ATTAAAGAAA ATTTACGAGA AGTTTCTGGA CAACCTTATG TTGTAGAAAA	4200
	TGGCTATTTA TFCGTCTTTG TTATTGATTA TTATCGTCAT CATTTAGTTG ATCAACATGC	4260
45	TGAAACTGAT ATGGAAAATG CATATGGTTC AACGGAAGGT TTGCTAGTAG GTGCAATCGA	4320
	TGCAGCATTA GTTGCCGAAA ATATTGCGGT AACTGCTGAA GATATGGGGT ATGGCATTGT	4380
	CTTTTTAGGA TCATTAAGAA ATGATGTTGA ACGCGTTCGA GAAATTTTAG ACTTACCTGA	4440
50	CTATGTCTTC CCGGTATTTG GTATGGCAGT AGGGGAACCC GCAGATGACG AAAATGGTGC	4500
	AGCCAAGCCA CGCTTACCAT TTGACCATGT CTTCCATCAT AATAAGTATC ATGCTGATAA	4560
55	GGAAACACAG TATGCACAAA TGGCAGATTA CGACCAGACA ATCAGCGAGT ACTATGATCA	4620

	CAAAGCAAGA TTAGATATGT TAGAACAATT GCAAAAATCA GGCTTAATAC AGCGATAgCA	4740
	AGATACCAAAA ATAACCCGCC CCCCTCTAGC TTAAATGAT AAGTATAGCT AGAGGGGGCG	4800
5	GGTATTTCTT GCAATGAATT AGTGTGAAGT TAATGCAGCA TTATCATTTG AATCGAAAGT	4860
	ATCTTTATCC CAATGTTTAG TTAACCTGGC GGTACCTGTA CCAGCTAGCA TTGAATCGTT	4920
10	CACGTTTAAT GCTGTTCTAC CCATGTCAAT CAATGGTTCA ACGGAGATGA GCACGCCGGc	4980
	TAAAGCGACT GGCAAGTTTA ACGTTGACAA CACCAATATG GATGCAAATG TAGCCCCGCC	5040
	ACCGACGCCA GCAACGCCGA ATGAACTAAT AATCACGACA GCGATTAACG TTACAATAAA	5100
15	TTGTAAATCA ATTTCTACAT TAGCGACGGG TCGGACCATA ATTGCAAGCA TGGCAGGGTA	5160
	AATGCCTGCA CAACCATTTT GTCCAATCGA CAATCCAAAT GTCGCAGCGA AATTGGCAAT	5220
	ACCTTCTGGC ACGCCTAGAC GTCTTGTTTG TGTTTGTACA TTCAATGGTA AGGCACCCGC	5280
20	GCTTGAGCGT GATGTGAATG CAAAGATTAA TACTTCCAAA GTCTTTTAA CATAGCGAAT	5340
	TGGGCTAATA CCTAACAGGC TTAAATAAT TAAGTGAATG ATATACATCG TAATTAATGC	5400
	AGCGTACGAT GCGATTAAGA ATTTTCCTAA AGTCCAAATG GCGCCAAAGT CACTTGTCGA	5460
25	TAATGTGTTG GCCATAATTG CTAATACACC GTATGGCGTT AAACGTAAGA CGAACGTCAC	5520
	AATCGCCATT ACTAGTGAAT AGATAGCGTC AATCGCACGC TTAAGCAATT CACCATGATC	5580
30	AGGTGTGTTG CGTnTACGCG TAAATAAGCA AATCCTATAA ACGAAGCAAA TATCACGACA	5640
	GCAATCGTGG aAGTTGCACG TTGTCCaGTG AAATCTAAGA ATGGATTTTT AGGCAATAAT	5700
	TCCAAAATTT GTTGTGGTAA CGTATGTGCT GTTAAATCTT TCGCTTGTTT AGCAATTTTCG	5760
35	CTTCCACGTG CTTGTTCAGC GTTACCAAGG TTAATTGTTG ATGCATCTAA ACCAAACACC	5820
	AAGGCATACA CAACACCAAC AATCGCAGCA ATGGTGACAG TGCCAATTAA AAAGATAAAA	5880
	ATGASACTAC CAATTTTAGC AAACTTTCT CCGATTTGAA TTTTAGTGAA TGCAGCTACA	5940
40	ATAGAAATGA AAATTAAAGG CATAACAATC ATTTGCAACA ATGCAACGTA ACCTTGTCGG	6000
	ACAATGTTGA ACCAGTCACT TGTGATGTA ATAACATTG AATGTGTGCC ATAAATAAGA	6060
45	TGCAATAACA CACCGAATAC TATACCAATC CCTAAAGCTG TAAACACACG TTTCGCAAAA	6120
	GATATATGTT TGCAGCCAT CATGTGCAAT ATTACGATGA AAATCACCAA TACAATAATA	6180
	TTAATCAGTG TAAGAAAAGC ATTCATGAAC GTCACCTCTT AAATTTTTGA ATATAATTCC	6240
50	GACTAGTATG CT	6252

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6730 base pairs

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

	ATCAAATCnC AAAATATTTA TTAATnAnAA GGGGATTATC CaTGTgAGAA ACAAAGTAAT	60
10	GCTCTTTTTT TACCTCTTGT GGGTTGAAAA aTGGATCATC AGAGATAGAC TTCTTCTTTT	120
	TCGAAGATGA CATTTGATAC TTTAATCTTC TAAAACCATA ACTTGTCGCA TCAAAAATGC	180
	CTTCTGTGAC AAGTAAAATC AAAAATATGC TAATAAAAAT AATTAATGAA ACATAAAACA	240
15	ATATATTTAA ATATGTAATG ATAGTATGGC TATTA AAAAAG CCATATAATA AACGTTAATA	300
	TTGGCGTTAT TAGTGCCATT CCAAGCCATT TTTTCAACAT TTGATCACTC CCACTTATAG	360
	AAAACCTCTTA CGCATAGTTT ACATTAAAAT CAGACATTGA GGAATGATTT TTTAATTTCT	420
20	TCAGCTTTAT TGAAATTCTA AAATCAATCA TTCTTCATTA GTTTAAAGCA AAAAAATATT	480
	GATATATAGT AAATATTGTA TATATAATAT TAGTTAAGAT TTCaGAAAAT TTTGAAGGGA	540
25	ATGGAAATTT AGAAATCGGA ATTTGTTAGA GGAGGGGATT AGATGGGGAA ATATATTTTC	600
	AAACGATTTA TTTATATGCT TATTTCTTTA TTTATTATTA TTACAATTAC ATTTTCTTTA	660
	ATGAAATTAA TGCCAGGTTC GCCATTTAAC GATGCTAAAT TAAATGCTGA ACAAAAAGAA	720
30	ATTTTAAATG AAAAAATATGG ATTAAATGAT CCTGtAGCTA CGCagTATTT ACATTATTTA	780
	AAAAATGTTG TTACAGGCGA TTTTGTAAT TCATTCCAGT ATCATAATCA ACCTGTGTGG	840
	GATTTGATTA AACCGAGACT ACTACCTTCT TTTGAAATGG GTCTTACAGC AATGTTCaTC	900
35	GGTGTGATAC TGGGACTTAT TTTAGGTGTT GCAGCAGCTA CTAAACAAAA TTCTTGGGTT	960
	GACTATACAA CTACAGTTAT TTCAGTTATT GCAGTATCTG TACCATCTTT TGTACTTGCT	1020
40	GTACTTTTAC AATATGTATT TGCAGTTAAA TTAAGATGGT TCCCAGTAGC TGGATGGGAA	1080
	GGTTTTTCGA CCGCGGTATT ACCGTCACTT GCATTATCTG CAGCTGTTTT AGCAACTGTC	1140
	GCCAGATACA TAAGAGCAGA GATGATAGAG GTATTAAAGT CAGACTATAT TTTATTAGCG	1200
45	AGAGCTAAAG GTAATTCGAC AATGCGTGTA CTTTTTGGAC ATGCACTTAG AAATGCTTTA	1260
	ATTCCAATTA TTACAATTAT CGTCCCATG TTAGCAAGTA TTTTAACAGG CACTTTAACA	1320
	ATTGAAAATA TTTTGGAGT TCCTGGATTA GGGGATCAAT TCGTACGTTC AATTACAACA	1380
50	AATGATTTCT CAGTAATCAT GGCAATCACA CTATTATTTA GCACACTGTT TATCGTTTCT	1440
	ATTTTTATTG TAGATATTTT GTACGGTGTG ATAGATCCAC GAATTCGTGT TCcAAGgAGG	1500
55	TAAAAAATAA TGGCTGAAAA TAAAAACAAT TTGTCGATTA ACGACGATCA TTCTAATGCA	1560

	TGAATCAGGA ACCTGAAATG CAACGAGAAA GCAAAAACTT TTGGCAAGAT GCTTGGGCTC	1680
5	AGTTAAAACG AAATAAGTTA GCTGTTGTCG GTATGATAGG TTTAATTATC ATTGTAATAT	1740
	TTGCTTTTAT CGGTCCAGTT ATAAATAAAC ATGATTATGC TGAACAAAAT GTAGAACATA	1800
	GAAATCTTCC GGCAAAAATA CCTGTATTAG ACAAAGTTCC ATTTTACCT TTTGATGGTA	1860
10	AAGATGCAGA TGGCAAGGAT GCTTATAAAG CAGCAAATGC TAAAGAAAAT TATTGGTTTG	1920
	GTACTGATCA GTTGGGTCGA GATTTATGGA CAAGAACATG GAAAGGTGCT CAAATTTTCAT	1980
	TGTTTATCGG TGTTGTTGCA GCGATGTTAG ATATTTTTAT TGGTGTGTA TATGGTGCGA	2040
15	TTTCTGGATT CTTCCGGTGA CGTGTCGATA CGATTATGCA ACGTATACTT GAAGTCATAG	2100
	CATCTATTCC GAATTTAATT GTCGTAATTT TATTTGTATT AATTTTGA CCATCCATTT	2160
	GGACAATTAT ATTGGCTATG TCTATCACAG GCTGGTTAGG CATGAGCAGA GTTGTACGTG	2220
20	GAGAATTTTT AAAATTAAAA AATCAAGAGT TTGTCATGGC TTCGAAAACA TTGGGGGCTT	2280
	CAAAATTCAA ATTGATATTT AAGCATATTT TACCTAATAC ATTAGGTGCT ATCGTGGTTA	2340
25	CATCAATGTT TACAGTACCT AGTGCTATTT TCTTCGAAGC ATTTTAAAGT TTCATTGGTA	2400
	TAGGTGTACC CGCACCTCAA ACATCGTTAG GGTCATTAGT AAATGATGGG CGCGCAATGT	2460
	TATTAATTTA TCCACATGAA TTATTTATAC CAGCAATGAT TTTAAGTTTA TTAATTCIAT	2520
30	TCTTTTACTT ATTTAGTGAT GGATTACGTG ATGCATTTGA TCCGAAAATG CGTAAATAAA	2580
	AAGGGGGCAT AGCATATGAC TGAAAGAATA TTAGAAGTAA ATGATTTGCA TGTTTCCTTT	2640
	GATATTACAG CAGGGGAAGT GCAGGCAGTG AGAGGCGTAG ATTTTATTTT GAACAAAGGG	2700
35	GAAACATTGG CAATTGTTGG TGAATCAGGT TCAGGTAAAT CTGTAACAAC AAAAGCAATT	2760
	ACAAAATTAT TCCAAGGGGA CACAGGAAGA ATTAAAAAGG GAGAAATTTT ATTTTtaggg	2820
	GAAGATTTAG CAAAAAACC TGAAAATGAG TTGATTAAAT TACGTGGCAA AGATATTTCA	2880
40	ATGATCTTTC AAGATCCAAT GACATCTTTA AACCCAACGA TGCAAATTGG TAAACAAGTC	2940
	ATGGAACCAT TAATTAAGCA CAAAATTAT AGTAAAGCAC AAGCTAAAAA GCGCGCATTG	3000
45	GAAATACTAA ATCTTGTAGG TTTACCAAAT GCAGAAAAAA GATTTAAAGC ATATCCTCAT	3060
	CAATTTTCAG GTGGACAAAG GCAAAGAATT GTTATTGCAA CCGCATTAGC TTGTGAACCT	3120
	AAAGTGCTCA TTGCTGATGA ACCAACGACT GCATTAGACG TAACGATGCA GGCACAAATT	3180
50	TTAGATTTAA TGAAAGAACT ACAACAAAA ATCGATACAG CAATTATTTT TATAACGCAT	3240
	GATTTAGGGG TTGTTGCGAA TATTGCTGAT AGAGTGGCAG TTATGTATGG TGGTCAAATG	3300
55	GTTGAAACAG GAGATGTTAA CGAAATATTT TATGATCCAA AGCATCCATA TACATGGGGA	3360

	GGAGCGCCAC CTGATTTATT ACACCCACCT AAAGGTGATG CATTGCGAG ACGTAGcAAT	3480
5	ATGCATTAGA TATTGATTTT AAAGTAGAAC CACCGTGGTT TAAAGTTTCA CCGACACATT	3540
	TTGTGAAATC TTGGTTATTA GACGCACGTG CACCAAAAGT TGAAC TACCC GAGCTGGTAA	3600
	AACAACGTAT GAAACCGATG CCTAATAATT ATGAAAAACC ACTCAAGGTA GAAAGGGTGT	3660
10	CGTTCAATGA AAAATGATGA AGTGCTATTA TCTATTAAAA ATTTAAAGCA ATATTTTAAC	3720
	GCAGGAAAGA AAAACGAAGT GgaGCGATTG AAAATATTTT GTTTGATATA TACAAAGGGG	3780
	AAACATTAGG TTTAGTAGGA GAATCGGGGT GTGGTAAATC TACAAC TGGT AAATCAATTA	3840
15	TTAAACTTAA TGATATTACA AGTGGAGAAA TTTTGTATGA GGGTATTGAT ATACAAAAGA	3900
	TTCGTAAACG TAAAGATTG CTTAAATTTA ATAAAAAGAT ACAGATGATT TTTCAAGACC	3960
	CATATGCGTC TTAAATCCT AGGTTAAAAG TAATGGATAT AGTAGCTGAA GGTATTGATA	4020
20	TCCATCATTT AGCAACTGaT AAGCGTGACC GAAAAAACG TGTCTATGaT TTA CTTGaAA	4080
	CTGTTGGATT AAGTAAAGAA CATGCCAATC GCTATCCTCA TGAATTTTCA GGTGGaCAAC	4140
25	GCCAACGTAT TGGaATTGCC CGTGcATTAG CCGTTGaACC AGAATTCATT ATCGCGGACG	4200
	AACCAATATC GGCATTGGAT GTTTCAATCC AAGCTCAAGT AGTTAATTTA TTATTAAAAT	4260
	TACAACGTGA AAGAGGGATT ACGTTCCCTAT TTATAGCTCA TGATCTATCA ATGGTGAAGT	4320
30	ATATTTTCAGA TCGTATTGCA GTCATGCATT TTGGGAAAAT AGTTGAAATT GGACCGGCAG	4380
	AAGAAATTTA TCAAAATCCA TTACACGATT ATACTAAGTC TTTATTATCA GCCATTCCAC	4440
	AACCTGATCC TGAATCAGAA CGCAGTCGCA AACGATTTAG TTATATTGAT GATGAAGCAA	4500
35	ATAATCATTT AAGACAATTA CATGAAATTA GACCGAATCA CTTTGTCTTT AGTACTGAAG	4560
	AAGAAGCGGC ACAACTACGA GAAAATAAAT TGGTGACACA AAATTAAGGG GAAGGGGGAA	4620
	ATGcAATGAC GAGAAAATTT AGAACACTTA TTTTAATTTT GATTGCTACA ATTGCATTAA	4680
40	GTGGTTGTGC TAATGACGAT GGTATTTATT CAGATAAAGG TCAAGTATTC AGAAAAATTT	4740
	TGTCATCAGA CTTAACATCC CTTGATACAT CATTATAAC GGATGAAATA TCTTCTGAAG	4800
45	TGAcTGCGCA AACATTCGAA GGTTTATACA CATTAGGAAA AGGTGACAAA CCGGTGTTAG	4860
	GTGTTGCGAA AGCTTTTCCT GAAAAGAGTA AAGATGGTAA AACTTTAAAG GTTAAATTAA	4920
	GAAGCGATGC TAAATGGAGC AATGGTGACA AAGTGA CTGC ACAAGACTTT GTTTATGCTT	4980
50	GGAGAAAAAC AGTTGACCCT AAAACAGGTT CTGAATTTGC ATACATTATG GGGGACATTA	5040
	AAAATGCGAG TGATATTAGT ACTGGTAAGA AACCTGTAGA GCAATTAGGT ATCAAAGCAT	5100
55	TAAATGATGA AACATTACAA ATTGAATTAG AAAAGCCGGT TCCATATATT AATCAATTAT	5160

ACGGTACGGC AGCTGATAGA GCGGTATACA ATGGTCCaTT TAAAGTTGAT GATTGGAAAC 5280
 AAGAAGATAA AACCTTACTA TCTAAAAATC AGTATTATTG GGATAAAAAG AATGTAAAAT 5340
 5 TAGATAAAGT GAATTATAAA GTTATTAAAG ACTTACAAGC CGGTGCAATCA TTGTATGATA 5400
 CTGAATCAGT AGATGACGCA TTTATTACTG CAGATCAAGT AAATAAATAT AAAGACAACA 5460
 10 AAGGATTAAA CTTTGTGTTA ACGACTGGGA CATTTTTTGT AAAAATGAAT GAAAAACAAT 5520
 ATCCTGATTT TAAAAACAAA AATTTAAGAT TGsTATCGCA CAAGCAATAG ATAAAAAAGG 5580
 ATACGTTGAT TCAGTGAAAA ACAATGGCTC AATTCCTTCC GATACACTAA CAGCCAAAGG 5640
 15 AATTGCGAAA GCGCCTAATG GCAAAGATTA TGCGAGTACC ATGAATTCGC CTTTAAATA 5700
 TAATCCTAAA GAAGCAAGAG CACACTGGGA CAAAGCTAAA AAAGAGTTAG GTAAAAATGA 5760
 AGTGACATTT TCAATGAACA CAGAAGATAC ACCAGATGCA AAAATATCTG CTGAATATAT 5820
 20 CAAATCGCAA GTTGAGAAAA ATTTACCAGG AGTTACTTTG AAAATTAAGC AATTACCGTT 5880
 TAAACAAAGA GTATCACTAG AACTGAGTAA CAATTTTGAA GCATCACTTA GTGGTTGGTC 5940
 TGCAGATTAC CCTGATCCTA TGGCTTATTT AGAAACAATG ACCACAGGTA GCGCACAAAA 6000
 25 TAATACAGAC TGGGGTAATA AAGAATATGA TCAATTACTT AAAGTAGCAA GAACCAAATT 6060
 GGCACCTCAA CCGAACGAAC GATATGAAAA CTTGAAAAAA GCAGAAGAAA TGTTCTTAGG 6120
 30 AGATGCACCG GTAGACCAA TTTATCAAAA AGGTGTtGCA CATTTaACAA aTCTCAAGT 6180
 AAAAGGATTA ATTtACCATA AATTTGGTCC AAATAACTCA CTAAACATG TATATATTGA 6240
 TAAATCGATA GATAAAGAAA CAGGTAAGAA GAAAAAATAA TATGCTTTGT AAATTAGGCT 6300
 35 GGAGACATAT CTCCAGTCTT TTTGTGTTGG ATAAAAaCTT TGGAATAAAA AATTTAAAAT 6360
 AAGTCGTTTT TTAAATTACT GAAATTGATT AAATGCATAA ATAAGTGAAT ATTCTAAAAA 6420
 TAAAcTTGTA ATAATTTTTT CTATGAGTAA ACTAAAAAGA AAAAATTAGA TTGAAAGTAG 6480
 40 GAGGCATATG TATGGGGAAG CTAATTAAAT ATATTTCAAT ACTTCTTATT GTCGTTTTAG 6540
 TGTGAGTGC TTGCGGAAAA AGCAGTAATA AAGATGAAGG AGTAAAAGAT GCTACTAAAA 6600
 45 CGGAAACCTC AAAACATAAA GGTGGTACCT TAAATGTAGC ATTAACAGCA CCGCCAAGTG 6660
 GTGTTTATTC TTCGTTATTA AATAGTACAC ATGCAGATTC TGTAGTTGAG GGATATTTTA 6720
 ACGAAAGCTT 6730

50 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6482 base pairs

(B) TYPE: nucleic acid

55 (C) STRANDEDNESS: double

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	AATTTTGTGTC ATTATTAAAA ACCTCGCTTT TAAAAGATTG AAAAGTAAAT GAGTGAAATT	60
	AAAGATTATG CACATTAAAA TCACGCCACA ATTTAATTGT GAAAAATATC ACAAATATAT	120
	TATAACACTA AATTTCCCAA AATTCAAAAG TGTGTTTTAT TGCAGAAAAC TTATAACaYg	180
10	TGCACAAGTT ATAGTGAATT GCAAACGGAT TACTTTAGTC TTTTAAAAAC ATGAAGTATA	240
	ATTTGTATAG CAATAAATAT AAAAATGGGA GGCTATGTTC AATGAGCAAT ATGAATCAAA	300
	CAATTATGGA TGCATTTTCAT TTCAGACATG CGACTAAGCA ATTCGATCCA CAAAAGAAAAG	360
15	TTTCGAAAGA AGATTTTGAA ACAATATTAG AGTCAGGTAG ATTGTCTCCA AGTTCTCTTG	420
	GGTTAGAACC TTGGAAGTTT GTCGTGATTC AAGATCAAGC GTTACGTGAT GAATTAAAAG	480
	CGCACAGTTG GGGCGCAGCA AAACAATTAG ATACAGCGAG CCATTTTGTG CTAATTTTGTG	540
20	CGCGTAAAAA TGTAACGTCA AGATCACCGT ATGTACAACA TATGTTAAGA GATATTAAAA	600
	AATATGAGGC ACAAACGATT CCAGCTGTTG AACAAAAATT CGATGCATTC CAAGCAGATT	660
	TCCATATTTT TGATAATGAT CAAGCCTTGT ATGACTGGTC AAGTAAACAA ACGTATATTG	720
25	CATTAGGCAA TATGATGACG ACAGCCGCAT TGTTAGGTAT TGATTCATGT CCGATGGAAG	780
	GTTTTAGTCT GGATACAGTG ACAGACATTT TAGCAAATAA AGGGATCTTA GATACTGAGC	840
30	AATTTGGTTT ATCAGTGATG GTCGCATTTG GCTACAGACA ACAAGAGCCA CCGAAAAATA	900
	AAACACGCCA AGCTTATGAA GATGTTATTG AATGGGTTGG ACCAAAAGAA TAAATAGAAT	960
	ACCGTATGTC TAAATATATA AAATTAAAAA GTTAGCAATA AAAAAGCCTG CGATTACATA	1020
35	AATGAATCGC AGGCTTTTGC GTGAAAAAAT TGTATTAATA AAGTATGGAT GATTATTTTT	1080
	CTGGACAAG GTCAGTATTT GAATGAACTG TGATGTCAAA CCCTTCTGGT GCCGTAAATG	1140
	TATGTGTTGA GCGTCGGGT TGATAAATAT CAACATGTGT TAATCCATAA CTTTGTGAAT	1200
40	TGTTTTGTCT TGCTTGATTG GATTGCCAAG TATTAGCAGC AATATGATGG TGATAATGAT	1260
	TCGTTGACAT AAATAGCGCA CGTGGAAAAT CAGACACATG TTGGAATCCT AATTGTTCAA	1320
45	TGTAACATTG ATATGCTGCG TCTAAATCAT GTGTTTTTAA ATGTAAGTGT CCAATCATGC	1380
	CTTTTGCTGG CATTCCTTGC CAACCTTCAT CAGTACGATG TGTTAATAAG GTTTGGCTAT	1440
	CAACTTCTAA AGTATCCATT TTAACCTTGC CATTTTGCCA TTCCCATGAA GATGAAGGTC	1500
50	TATCGCGATA GACTTCAATA CCATTACCTT CGGGGTCGTT GAAATATAAA GCTTCACTTA	1560
	CTAAATGATC ACCAGCGCCG ATGCCCATAT TTTTTGTGC CACGAAATAT AAGAAGTTAG	1620

55

	aAGTCTGACG GcCGTCTTCT AATAAATGTA ACGTTAGAGT ATGGcCACCA GTCCCAACAG	1740
	ATAATACGGT TGTATTATCG TCAGAACTTT TAACGGATAG TCCTAAAATG TTTTGTAAA	1800
5	ATGTTGTCAT TAAGTCTAAG TCTCTTACGT TCAGTACAAT GTTTGTCACT TGTGTTGCTG	1860
	TTTTATCGTG AAATGCCATT ATGCATCGCC TCTTTTCTA TTTTCTATA AGTTAGTATA	1920
10	AAAAGTATAC CAGAAAAGAA AATGAATTGA TAGCATAAAG TTTGAAATGC AAAATAACTA	1980
	GTCGTTTTGC AATTTTATcAT TGATGCGAAC AAAAAAGCGA TGGTACAGTT GCACCATCGC	2040
	AAAATTTATT TAACCAAGAT ATACATCTTG ATATGAATCT TCTTTTCTA ACATATGTTT	2100
15	GGCAATGAA CATGAGGCAA TAATTTTCAA ATTATTTTCT CGAGCGTGTT CAACAACTGc	2160
	TTTAAGTAGT TTTTGGCCAA CACCTTGACC ACCAAGTTCA TCAGATACGC CTGTATGATC	2220
	AATGTTAATT TCATTATTAT CCACAAAACG GTATGTGATT TCAGCTAAAG CATTATTTTC	2280
20	ATCATCACCA ATATAGAATT TGTTCTCGCC TTGTTTGATT TCAAGGTTAC TCATACATAT	2340
	CAACTCCTAT CATGATTGAT TATAGTATTT CCCTATTCTA TTTTAACTTA AACGAAGTCA	2400
	AAGGTGCATG ACAGTCATGT GACGACATTG CCACATCTAT GTAGTCGTTT TTATTAAGCA	2460
25	CAGTTTGAAA TGAAGATGAA AACACGTATC TTGACATTAA ATCTATTCAG CTATATAATT	2520
	TATCTCGAAA TCGAAATAAA ATAAAAAAGT TGGTGATCAT ATGGATCGAA CGAAACAATC	2580
30	TCTCAATGTT TTTGTGGAA TGAATAGGGC GTTAGACACA TTAGAGCAAA TTACAAAAGA	2640
	AGACGTAAAG CGATATGGCT TAAATATTAC TGAATTTGCA GTGCTCGAGT TGCTTTATAA	2700
	TAAAGGTCCG CAACCAATTC AACGTATTAG AGACCGCGTA TTAATTGCAA GTAGCAGCAT	2760
35	TTCATATGTT GTAAGTCAAT TAGAGGACAA AGGTTGGATT ACACGTGAAA AGGATAAAGA	2820
	TGATAAACGT GTATATATGG CTGTGTTAAC TGA AAAAGGT CAAAGTCAAA TGGCAGATAT	2880
	TTTCcCTAAG CATGCTGAGA CATTAAACAA AGCGTTTGAT GTGTTAACAA AGGATGAATT	2940
40	AACAATCTTA CAACAAGCGT TTAAGAACT AAGTGCACAA TCTACAGAAG TGTAAGGCGT	3000
	GCACTAAAAA TTTACATTAA AGTATCTCGA TTTCGAGATA AATGCACTAA AAATATAAAG	3060
	AGGGTATATA AAATGATAAA TAATCATGAA TTACTAGGTA TTCACCATGT TACTGCAATG	3120
45	ACAGATGATG CAGAACGTAA TTATAAATTT TTTACAGAAG TACTAGGCAT GCGTTTAGTT	3180
	AAAAAGACAG TCAATCAAGA TGATATTTAT ACGTATCATA CTTTTTTTGC AGATGATGTA	3240
50	GGTTCGGCAG GTACAGACAT GACGTTCTTT GATTTTCCAA ATATTACAAA AGGGCAGGCA	3300
	GGAACAAATT CCATTACAAG ACCGTCTTTT AGAGTGCCTA ACGATGACGC ATTAACATAT	3360
	TATGAACAGC GCTTTGATGA GTTTGGTGTT AAACACGAAG GTATTCAAGA ATTATTTGGT	3420

55

	TTAAATGAAG GGGTAGCACC TGGTGTACCT TGGAAGAATG GACCGGTTCC AGTAGATAAA	3540
	GCGATTTATG GATTAGGCCC CATTGAAATT AAAGTAAGTT ATTTTGACGA CTTTAAAAAT	3600
5	ATTTTAGAGA CTGTTTACGG TATGACAACCT ATTGCGCATG AAGATAATGT CGCATTACTT	3660
	GAAGTTGGCG AAGGAGGCAA TGGTGGCCAG GTAATCTTAA TAAAAGATGA TAAAGGGCCa	3720
10	GCaGCACGTC AAGGTTATGG tGAGGTACAT CATGTGTCAT TTCGTGTGAA AGATCATGAT	3780
	GCAATAGAAG CGTGGGCAAC GAAATATAAA GAGGTAGGTA TTAATAACTC AGGCATCGTT	3840
	AATCGTTTCT ATTTTGAAGC ATTATATGCA CGTGTGGGCG ATATTTTAAAT AGAAATTTCA	3900
15	ACAGATGGAC CAGGATTTAT GGAAGATGAA CCTTATGAAA CATTAGGCGA AGGGTTATCC	3960
	TTACCACCAT TTTTAGAAAA TAAAAGAGAA TATATTGAAT CGGAAGTTAG ACCTTTTAAAT	4020
	ACGAAGCGTC AACATGGTTA ATTGGAATGA GGAGGATTG TGATGGAACA TATTTTTAGA	4080
20	GAAGGACAAA ATGGTGCGCC AACACTAATA TTATTGCATG GTACAGGTGG TGATGAGTTC	4140
	GATTTATTAC CGTTAGGCGA AgcATTGAAT GAAAATTATC ACTTGTTAAG TATTAGAGGA	4200
	CAAGTTTCAG AAAATGGGAT GAACCGTTAT TTCAAACGTC TTGGTGAAGG TGTTTATGAT	4260
25	GAAGAAGATT TGGCATTTCG TGGACAAGAA TTGTTGACGT TCATTAAAGA AGCTGCTGaA	4320
	CGTTATGATT TTGaTATTGA AAAAGCAGTA CTTGTTGGAT TTTCAAATGG ATCAAATATA	4380
30	GCGATTAACT TAATGTTGCG TTCAGAAGCA CCATTTAAAA AAGCATTGTT ATATGCACCG	4440
	TTATACCCAG TTGAAGTAAC GTCAACAAAG GATTTATCAG ATGTCAGTGT GTTGCTTTCT	4500
	ATGGGGAAAC ATGATCCAAT TGTGCCATTA GCTGCAAGTG AACAAGTCAT TAACTTGTTT	4560
35	AATACACGTG GGGCACAAGT CGAAGAAGTT TGGGTGAAGG GCCATGAAAT TACAGAACT	4620
	GGATTAAACG CTGGTCAACA AATACTTGGG AAATAACAGT TCTATTAAGA AGCGGACAGA	4680
	TGGA ¹ AAAGAT TTTTACTTTT CATCTGCCCC CTTTTTTGAT TTTGAAGTGC TGTAATAAAT	4740
40	TTTACAATAG TATAGATATT TTAATCGATA TGAGATTTGC CGGTAATACG CTTAATTAAA	4800
	CCTTTATAGA GTACAGGTAT GAGTAAGATG AAACCGAACA ATCCCATAAT AGGGAATACT	4860
	TTTCCAATTA ATGAAATGAa ACCGATAAAT GACTAATAT AAGTGATGAC AGCCATTGTA	4920
45	ATAATAATGA TGAAGTAACG TCTGCTGAAT GGAACGCTGA AACGTGACGC AAATGCATAC	4980
	ATTAATCCAA CAACAGTATT GTAGATGACA AGTATCATAA TGACAGACAT AATAATACCA	5040
50	ATTGACGGAG ACATTTGTGT CGCTAATTTT AATGTAGGTA GATCTACGTG TTTAATTTTA	5100
	TCGAATTGAG AAATTAAACC TAGATTAATC ATCATGAGTA AAAATGTAAT GATTAAACCG	5160
	CCAATCAAGC CCCCCTATAA CGTTGAGTCA CGATATTTAA CTTTACTACC CATCACTGAT	5220

55

CCAGGTGATA ATGATTTCTG CTTATGAATC TGAGCATCAT TATTAGCGGC AGTAAAATCA 5340
 AGATGACTTG TTGTGAAATA GTAGACCGCA ATCATAATGA CAATCGCAAT TAAAAATGGG 5400
 5 GTAACACCGC CAAGCACAGC AATTAAACGA TCGAATTTTA GAAACAGTGT TGCTAAAATA 5460
 AAGGCGACTA ATATGAGTGC GCTCAGCCAA TACGGTAAGT TGAAACTTTG ATGAATGGTT 5520
 10 GACGCACCAC CTGCAGTCAT AATAATAGCT AAAGACAACA TAAACATTGT TAAAATAATA 5580
 TCAAAACCTC TTGCAATAGA GGGGTATAAG AAATAGTTAA TTGAATCAGA ATGATTTCTG 5640
 GACTTTAGAT GATGACCTGT ATGCATGACA ACCATTCCAC CTAAAGTAAT CAATAGTCCT 5700
 15 GTTACAATAA TGCCTGAAAT GCTATATGCG CCATGACTTG TGAAAAACTG GAAAATTTCT 5760
 TGACCAGTAG CAAAGCCGGC ACCAACGACA ACACCAACAA AGGCAAATGC CACAATAATG 5820
 GACTCTTTTA AGATACGCAT GATTTAAAAA TGTCCCTTCG TAATTTTAAG TAATATAGAA 5880
 20 AATGTAACAT ACATGTTAAT GAAAAATATA GTACTAATAT AGTATTTTGT TAAATTGGAG 5940
 TAGAAGCGAG GGTGTCGGTC ATTTCAATTA TTTATTAGTT GATTTTGCAT TTTTTTGCTG 6000
 TAAAGTTGTT ATAATACAGT TAACAGGAAT TAGCATAGAT ACACCAATCC CCTCACTACT 6060
 25 CGCAATAGTG AGGGGATTTT TTTCGGTGTA GCTAGGTCGC CTATTTATCA TCGTGTTCG 6120
 GTAGCaATGC GTAAACACAG TACCACTAAA TAAGTGCACG ATACATGCAT CAAATGTCGT 6180
 CTTTAGTcTA AGTAACGATC ATGCATTAAC ATTTTCAAAA TATCTATTG AGCTTGAAGA 6240
 30 TCTTTACCAA TATTGGTATC ACGAATCTTC TTACGTTGTA ATTCTTTATC TACGACCGC 6300
 TTTATAGAAA GTTCATCGAT ACCTTCGGAA AGTATTTTTn CTTTAGCGTT AAATTGTTGG 6360
 35 TGTGCAACGA GTTGCAATACC GAATGAATTA TACAATAGTG TATAGCCTGC AATGCCAGTn 6420
 GTTGACTION AAGCTTTTGA AAAGCCACCA TCAATGACAA GCATCTTTCC ATCAGCCTTG 6480
 AT - 6482

40 (2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16592 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ATTTAAGGCG ATTGCTTG TG TATTTCTCTC TTTTGTAGGC AAACCTGCAC TCGTTCCAAA 60
 AAATGTAAC TCCATATATG CCCCTCCTT TCTTCAATTC ATTTTATCAT AAAATTTGTA 120

55

	AATTTTCTA ACTTTAACGT AGACATAACT ATATAAATTT TGATAATTAC GTTATACTTA	240
	TCATTAATAA GTATCACATT AAACATGATA CATGAATCGA TATTTCATTT AAGACACTGC	300
5	ATACAGTCGA GCATATTGTA TGACCTACTG AATGGATTAT CTTATAATAA TAAATCATAT	360
	ATCTAATTAA GAATTGAGGT TTTAATCTTG AGTACTAAAA ACAAACACAT CCCATGTTTA	420
10	ATCACAATCT TTGGTGCACT GCGTGACTTA AGCCATCGTA AGTnGTTTCC ATCAATATTC	480
	CATCTCTACC AACAAAGACAA TTTAGATGAA CATATTGCCA TcATCgGTAT TGGACGTCGT	540
	GACATkwnTA ATGATGATTT CCGTAATCAA GTAAAAATCAT CAATTCAAAA GCACGTAAAA	600
15	GATACAAACA AAATTGACGC GTTTATGGAA CATGTCTTCT ATCATAGACA TGATGTTAGT	660
	AATGAAGAAA GCTATCAAGA ATTACTAGAT TTTAGTAATG AATTAGATAG CCAATTTGAA	720
	TTAAAAGGTA ATCGACTATT CTATTTAGCA ATGGCACCAC AATTCCTTGG CGTTATTTCT	780
20	GATTATCTAA AATCTTCTGG TCTTACTGAT ACAAAGGAT TTAAACGCCT TGTTCGAA	840
	AAACCATTCTG GTAGTGATTT AAAATCAGCC GAAGCATTAA ACAATCAAAT TCGTAAATCA	900
	TTTAAAGAAG AAGAAATTTA TCGTATTGAC CACTATTTAG GAAAAGACAT GGTTCAAAAT	960
25	ATCGAGGTAT TACGTTTTGC GAATGCGATG TTTGAACCAT TATGGAATAA CAAATATATT	1020
	TCAAACATCC AAGTTACATC TTCTGAAATA CTAGGTGTTG AAGATCGTGG TGGTTATTAT	1080
30	GAATCAAGTG GCGCGCTAAA AGATATGGTG CAAAACCACA TGTTACAAAT GGTTGcATTA	1140
	TTAGCTATGG AAGCACCTAT TAGTTTTAAAT AGTGAAGATA TCCGTGCTGA GAAAGTAAAA	1200
	GTACTTAAAT CACTGCGTCA TTTCCAATCT GAAGATGTTA AAAAGAACTT TGTTCTGGT	1260
35	CAATATGGCG AAGGCTATAT CGATGGTAAA CAAGTTAAAG CATACCGTGA TGAAGATCGC	1320
	GTTGCAGATG ACTCTAACAC ACCTACCTTT GTTTCAGGTA AATTAACAAT TGATAACTTT	1380
	AGATGGGCTG GTGTACCATT CTATATTCGT ACTGGTAAAC GTATGAAATC TAAAACAATT	1440
40	CAAGTTGTCG TTGAATTTAA AGAAGTACCA ATGAACCTAT ACTATGgAAA CTGaTAAACT	1500
	GTTAGATTCA AACCTATTAG TAATCAATAT CCAACCTAAT GAAGGTGgTA TCTTTtACAT	1560
	CtAAATGcTA AGaAAAATAC ACAAGGTATC gAAACAGrAC CTGtCCmATT GtCTTACTCm	1620
45	ATGaGCGcTC aAGaTAAAT GaATACTGTA GATGCATATG AAAATCTATT ATTTGATTGT	1680
	CTTAAAGGTG ATGCCACTAA CTTACGCAC TGGGAAGAAT TAAaATCAAC ATGGAAATTT	1740
50	GTTGATGCAA TTCAAGATGA ATGGAATATG GTTGaTCCAG AATTCCTTAA CTATGAATCA	1800
	GGTACTAATG GTCCATTAGA AAGTGATTTA CTACTTGCTC GTGATGGTAA CCATTGGTGG	1860
	GGACGATATT CAATAATTGA ATTAAAACGC ACATGTTAAA CAAAAATAAA TGAGCGAATG	1920

55

	TATATTATGA AATTATATTT TACAATGCCC AAAACTATTT TAATAATCAT TGAACAAATG	2040
	GGTGTATAAT TTATAGAAAT AATGTAGAAT AAAAATAAAT GATTGAATTA ATTGGAGTGA	2100
5	AAGTTTTGGA CGTTATCAAG CAAATACAAC AGGCAATTGT TTATATTGAA GATCGTTTAT	2160
	TAGAGCCTTT CAATTTGCAA GAATTAAGTG ATTACGTTGG TCTTTCGCCA TACCATCTTG	2220
10	ATCAATCATT TAAATGATT GTCGGCTTAT CTCCAGAAGC TTATGCACGC GCGCGTAAAA	2280
	TGACACTCGC TGCAAATGAT GTGATTAATG GTGCTACACG ACTTGTAGAT ATCGCTAAAA	2340
	AATATCACTA TGCAAATTCA AATGATTTTG CAAATGATTT TAGTGATTTT CACGGCGTAT	2400
15	CACCTATTCA AGCCTCTACT AAAAAAGATG AATTACAAAT TCAAGAGCGA TTATATATCA	2460
	AATTATCAAC TACTGAGAGA GCACCTTATC CATAACAGATT AGAAGAGACA GATGATATTT	2520
	CATTGGTTGG ATATGCACGA TTTATAGACA CTAAGTATTT GTCACATCCT TTTAATGTTC	2580
20	CGGATTTTTT AGAAGACTTG CTCATTGATG GTAAAATTAA AGAGTTACGA CGATATAATG	2640
	ACGTTAGTCC ATTTGAACTA TTTGTTATTA GTTGTCTCT TGAATGTTT TTAGAAATAT	2700
	TTGTAGGTGT ACCAAGTGAA CGTTATCCTG CACACTTAGA AAGTCGATTT TTACCTGGCA	2760
25	AACATTGTGC GAAATTCAAT TTACAAGGTG AAATTGATTA TGCAACTAAT GAAGCTTGGT	2820
	ACTATATTGA ATCAAGTTTG CAGTTAACAT TGCCATATGA ACGAAATGAT TTATATGTTG	2880
30	AAGTGTAACC TCTCGATATT TCATTTAATG ACCCATTAC TAAAATTCAG CTTTGGATTG	2940
	CTGTAAACA GAGTCCTTAT GACGAAGATT AAATAATAAA AAACAAAGAA GCCCCCTAAT	3000
	ATATCTATAG GTCTACAAAT GGCCTTAGAT TCTATTAGGG GGCATATTAA TATGTTAATT	3060
35	TAGTTCGATA ACACATGCTT CATATGGACG TAACTGTTTT AAATTAACIT TGGCATCATA	3120
	ATTAAATAGC TTTACTTCTC CATGGCTTAA ATCAAATGGT ACAGTTAATT CTGCTTCGTG	3180
	GTTAGTAAGA TTACCTACAA TAAGAACTTG CTTTTCATTT AATGTTCTCG TGTACGCAA	3240
40	AACTTGTGAA TTTTCAGCAT CTACTAAATC AAATTGACCA TATACGTATA CATCATTAGA	3300
	CTTTCCTAAT TGAATTAAAT CTTTATAAAA TTGTAATACT GAATGCTCAT CTTCTAATTG	3360
	TTGTGCAACA TTGATAGTTT TATAATTCGG ATTCACGTTG AACCACGGTT CACCATTGTG	3420
45	AAATCCTCCA TTTAACGTAT CATCCATTG CATTGGTGTG CGAGAATTAT CTCGGTTCTC	3480
	ATCTTTATAT TTCGCAAGTA AAGCGTCTAC ATCTCCACCT TGAGCTTTCA CTATTTGATA	3540
50	GTCATTTTTA ACAGCAACAT CGTTAAACGT TTCAATACTT TCAAATGGAT AATTTCGTCAT	3600
	ACCAATTTCT TGACCTTGAT AAATGAATGG CGTACCTTGT TGCAAGAAAT AAACAGCTGC	3660
55	ATGACTTGTT GCTGATTCAT ACCAATACTT GTCATCGTCA CCCCACGTCG ATACACGTCG	3720

	CCATCTATTT AATACAGATT TATACGAATT TACATCAAAG TGAGAATCAC CACTATTCCA	3840
	CAGTCCCAAA TGTTCAAATT GGAATATCAT ATTAAATTTA CCATTTTCTT CCCCAGCCCA	3900
5	GTCATCAGCA TCATCAGGGC TTACACCATT CGCTTCACCA ACAGTCATAA TGTCATACTT	3960
	ACTTAATGAG CGATCTTTCA TCTCTTGTA CCAAGTTTGT ATACCTGGCT GATTCATATC	4020
	TACATCAAAT GCTGGGGCAT ATGTTTTACC CTCAGGTACA GGTAAGTCAC CCGCTTCAAA	4080
10	CGTCTTCTTA ATATGCGTAA TTGCATCTAC TCTAAATCCA TCAATGCCTT TATCAAACCA	4140
	CCAGTTCATC ATTTCAAATA CAGCATCTCT AACTTCCGGA TTACCCCAAT TCAAATCAGG	4200
15	TTGTTTTTTA CTGAATAAAT GGAAATAATA TTGCTCAGTA TTAGCATCAT ATTCCCATGT	4260
	AGATCCATT AATATACTTT CCCAGTTGTT AGGTTTCAGAG CCATCTGGCT TTGGATCTTG	4320
	CCAAATGTAC CAATCACGTT TGGGATTGTC TTTACTAGAT TTGGATTCTA TAAACCAAGG	4380
20	ATGTTTCATCA GATGTATGAT TTACAATAA ATCTAAAATA AGCTTCATGC CTCTATCATG	4440
	AACACCTTTT AATAAACGAT CAAAGTCTTC CATCGTTCCA AATTCATCCA TAATCTCTTG	4500
	GTAGTCACTA ATATCATAAC CATTGTCATC ATTAGGTGAT TTAAACATTG GACTGAGCCA	4560
25	AATGACATCG ATACCGAAAT CTTTAAAGTA GTCCAATTTA TCAATCATTC CAGGTAAATC	4620
	CCCAATACCA TCGTGATTAC TATCATTAAA ACTTCTTGGA TATACTTGAT ATGCTACTGC	4680
	TTCTTTCCAC CATGCTTAT TCATTTTAAA ACTCCTTTGC TATCGCTGTG TTGATTTTCT	4740
30	TATTTTTAAT TCTGTATCTA TAATGACGAG TTCAATAACA TCCTGTGCTT TGTTTTTCAA	4800
	TATATTTAAA ATTGCTGCAC CAGCCTGTTG ACCTAACATT CGAGGCTTGA TGTC AATACA	4860
35	GGTTTGTGGT GGTGACGCAA TTTCGGTTAA ATAAGAATCA TTGAACGTTG CTGTCATTAC	4920
	ATCTTTCCGA ATTTCAATAT TAAGTTCATA TAGGACACTT AAAATCGCTA AATGTAACAT	4980
	AGCATCTAAC GAAATGATTG CCTGTTTAAAT ATTTGGGTCC TTCAAACGCG TATGTAGATT	5040
40	TTGCATGTAA TTA AAAAATAA CTTCTCTTTC ATTACTAGTC TCAATAATTT GATAATTAAT	5100
	TTTATTTTGA GAAGCTATCG TTTCAAATCC TTGAATTCTA TCTTTTGAAA CTTCAA AATT	5160
	TCCTTTTCT GT AATAAATA TTAATTCATC TACACCTGT TCAATAACAT GTCGTGTCAA	5220
45	ATTTTCAGAA GCTAATATAT TATCATTATC TATATGTGTA AATTGATGAT CTATATCCGA	5280
	TGTAGGCTTA CCAATCACAA TAAATGGCAT GCTTTCATCA ATTAACATTT GTTTAATCGG	5340
	ATCATTTTCT TTTGAATAGA GCAGTATAAA CGCATCAACC ATTCGTTGTT TAATCATTTT	5400
50	ATAAACTTCA TCCATTAAAT CATTCAATT ATTGAGACT GTCGTTTGTG TACCATAGCC	5460
	ATGCTGGTTA CACGTTTCAG AAATTCCTAG CAATACATTG ATGTAGAATG GATTCAGTCG	5520
55		

	AGTTCTAGCA GCGGTATTAG GAAAATAAATT CAATTCTTCC ATAACCTTTCT TCACTTTTGA	5640
	AATTGTGCGCT TCGCTAATAC GTTGATTTC TTTTATAACT CTGAAACTG TCGAAGGAGA	5700
5	AACACCGGCT TTTAGTGCAA CATCTTTAAT CGTAACCATT TAATCACCTC CTGTTAATTT	5760
	CTGCATCGGA AAACGCTTCC AACCCTGTA TAATACCAGT TTAGTCACAC TTTCTAAAAA	5820
10	AGTCAAAAGA TTTGTGCAAA CGATTGCATA AAACGATAAA AATAAAACCT TCATACTGAA	5880
	ATTCAATCCG AAAATCAATA TAAAGGTTTG TATAAATATT AAAATCGATT GTTTAGTCAC	5940
	TAACTGCAAA ATAGTTACCT TGGCCATCTT GAAAATTAAA TACACGTTGA CCATTCAATTT	6000
15	CTACTATATC ATGCCCAGTT AAACCTAAAT CATTTAATTT TGAGTATAAT GCATCAAAGT	6060
	TTTTCTCTTT AAACATTAAA GATGGTGTTC CTAGGTTTAC TTCCGGGCTA TGCTTTTCAA	6120
	TAAATTCTTT TGCCATAATC GTCAATGACG TTTCAGCATC TTTGGTAGGT GATACTTCAA	6180
20	CTGCAACATA GTCCTCAGCT AACGGTGTTC CACTTACAAC AACAAATTCT AAAGTTTCTG	6240
	TCCAAAATGC TTTCGCTTTT TCGACATCAT CAACATATAA CATAACTTGA TTAACTTTT	6300
	CCATAAAATA GTACCTCTAT TTCTCTATAG TACATGCTAT CATAACACAG TAAATATTTT	6360
25	ATTACTTCAC AAAATGCTTA AAAATATGGC GGGATGCTTT TAAGGTCAAG GATAATACTT	6420
	GTGTAATTTT TTATAGGTTG TAGCTACTCT ATCACACTCT CTTTTATATT TATCAAAAGA	6480
30	TATAAAAAAG GATAGTATCT TTCAACTATC CTTTAATCAA TATTATTCTT CAATCCATTG	6540
	TGTATGGAAT ACGCCTTCTT TATCTTTTCT TTCGTACGTA TGAGCACCGA AGTAGTCACG	6600
	TTGTGCTTGA ATTAAGTTTG CAGGTAAATC AGCAGCACGG TAACTATCAT AGTAATTAAT	6660
35	ACTTGATGAG AAACCAGGTG TTGGTACACC ATTTTGAACA CCAGTTGCGA CAACATCACG	6720
	TAACGCATCT TGATATTCAG TAACGATGTT TTTAAAGTAA GGATCTAGCA ATAAGTTTTG	6780
	TAATCCTGGA TTATTATCGT AAGCATCTTT GATCTTTTGT AAGAATTGTG CACGGATAAT	6840
40	GCAACCTTCT CTCCAAATCA TAGCTAAATC ACCAAGTTTT AAATTCCATT CATTATCTTC	6900
	ACTTGCTTTA CGCATTTGcG CGAAACCTTG TGCATAAGAA CAAATTTTAC TCATATATAA	6960
	TGCTTTACGA ATTTTTTCTA AAAAGTCTTT CTGTGCACCA TCAAATGATG CTTTTGGACC	7020
45	ATTTAATTCT TTAGAAGCAT TTACGCGCTC TTCTTTGaTT GAAGAGATAA AACGTGCAAA	7080
	TACAGATTCA GTAATGATTG TTAATGGAAT ACCTAATTCT AATGCGTTAA TTGAAGTCCA	7140
	TTTTCCTGTA CCTTTTTGac CTGCAGTATC AAGAATTTTT TCAACTAATG CTTCTTTATT	7200
50	TTCATCTAAT TTCATGAAAA TATCACCAGT GATTTCATT AAATAACTTT CTAATTCACC	7260
	AGCATTCAG TCTTTGAACG TTTGAGCAAT GTCTTCATGA GACATGCCTA ATAATTCCTT	7320

55

	CATTTTCACA TAGTGTCCAG CACCATTAGG TCCAATATAA GTAACACATG AAGCACCGTC	7440
	TTTTGCCTTT GCAGCAATTG CATCAAGAAT ATCTGCAACT TTGTTATAAG CTTCTTCTTG	7500
5	TCCACCCGGC ATTAATGACG GACCAGTTAA CGCTCCAATT TCACCACCAG AAACGCCCAT	7560
	ACCAATAAAG TTGATTGCAC TTTGTGyWAA TGCTTTATTA CGTCTGATAG TATCTTGATA	7620
	GTTTGTATTA CCACCATCAA TTAATAATATC TCCATCATCT AATAAAGGTA ACAAACATATC	7680
10	AATCGTTGCG TCCGTAGCTT TACCTGCTTG AACCATTAAAT AAAATTTTAC GTGGTTTTTC	7740
	TAAAGAATTA ACAAATTCTT CCAATGAATA CGTTGGATGA ATATTTTTCC CTTTGTATTC	7800
15	TTCAACCATT AAATCAGTTT TTTCACTTGA GCGGTAAAT ACAGATACAC TATATCCGCG	7860
	TGATTCAATA TTCCAAGCTA GGTTTTTACC CATAACGGCT AAACCAATAA CTCCAATTTG	7920
	TTGTGTCATA TTACTTACCT CACTTGTTGA TTTTTCATTA GTATTGTATC ACAAATAGA	7980
20	CATACACTAC ACTAAATCAT TTCGAATGTC GCGCAACTAT TTTGATTATT TCTAACACTT	8040
	GACTTGCAAG CAAGTTC AAT GATTTAATCG GCATTCTCTC ATTTGTTGTA TGGATTTTTT	8100
	CATAACCAC TCCTAAAATG ACTGAAGGAA TACCAAATGT ATTAATAATA CTGCCGCTCTG	8160
25	AACCGCCACC AGAAATAATT GTATTTGCAG ATAATCCTAA ATTACGAGCA CTTTCTTGTC	8220
	CAATTTTAAC AACCGCTTCA TTATCATTAA TTTTAAATCC TGGATAACTT TGCTCCACTG	8280
	TAATACTGTC TTTCCACCT AATTCTGATG CAGTAGTTTC AAACACATCA GTCATATGTT	8340
30	TGACTTGTGT TTTTATTCTT TCTGGATCGT GAGAACGTGC CTCTGCTTCT AAAATGACTT	8400
	CATCTGCAAC AATATTGTA GCTGAACCGC CATGAACTT ACCAATATTG GCAGTAGTTA	8460
35	TTTCATCAAC TTGTCCTAAT TTCATTGAC TAATTGcTTT CGCCGCAATA TTAATAGCAC	8520
	TAACACCCTC TTTTGGCGTA CTTCATGAG CCGTTTTGCC AAAAATTTTA GCTGAAATTA	8580
	ACAATTGCGT CGGTGCACCT ACAACCGTAG TACCGACATC AGCACTTGCA TCAATAGCAT	8640
40	AACCAAAGTC CGCGTCCAAC AACTCTGAAT TTAATTCTTT AGCACCAATT AAACCTGATT	8700
	CTTCTCCAAC AGTAATCACA AATTGAATTT GTCCATGTGG GATTTGTTGT TCCTTTATCA	8760
	CTTGCAAAAC TTCAAGCATC GCTGATAATC CTGCTTTATC ATCTGCACCT AGAATAGTCG	8820
45	TACCATCAGA GTATATGTAG CCGTCATCTT TTACAATTGG CTTTACATTA ATTGCGGGTA	8880
	CAACAGTATC CATATGGCTC GTCAAATATA ATTTAGGTAC TTCGCCTTCT TCGATAGTAC	8940
	TATTCATTGT ACACACTAGA TTATTGGCAC CTAATTTAGG ATGTTTAGCC GCTTCATCTT	9000
50	CTTTAACATC TAACCCTAAT GCTATGAATT TTTCTTTTAA AATAGGTTGG ATTGTTGATT	9060
	CATTCCCTGT CTCAGAATCG ATTTGTACAA GTTCAAAAAA CGTATTAAGT AATCTTTGCT	9120
55		

	GATGAAATAA AATGTTACAG TAATTGACGT TACACAGATT TATCAGGTTT GTAAATTGTG	9240
	TCATATTATT TTCAATTTAT TATATATAAT TATTGTAACT CAACTAAGC TTTGTCAAAA	9300
5	ATATATTGAT TGATTTTTC AAGATATCGT ATAATGAGGA AAATGACATA AGCAAACCTTA	9360
	CTCATGTTTT TTATTATATT CCTTTATGAT GATTGCTAGT TATATCGTCT CAAGTTAAAA	9420
10	GTTTTATATC TTATGTCGTA ATTATTAATA CAAAGGTTAT TCATTGGAG GCACACAAAA	9480
	TGCAAATAA AGTTTTAAGA ATTATCATT TCGTTATGCT TGTATCAGTT GTATTAGCAT	9540
	TGTTATTAAC GAGTATCATT CCAATTTTAT AAACATATC TCACTACCT ATACAAAATC	9600
15	ATACAATTAA AAATCCATCC ATTATAAAG CATGTATTAA TAAGTTATCG TATTGCAACG	9660
	ATTACTTTCA AACATGGGTC ATACGGATGG ATTATTTTTT AAGCTACTTC ACTATGCATT	9720
	TTCAATGAAC CAAATTGCGA TTTGATTGT AAATATTCTT CTAATTCATT TAATATTTGA	9780
20	ATAATACTTG CTCTCGAGTT AAGCGCTTTG TGTGTGTTG GCAATGGCAG TTCATCCAAT	9840
	TTCAAACGCG TCTCATACAA ATTGTGTAAG CGCATTGCTG TATAGTCATT ACTATTCACA	9900
	TTTAGACCAA TTTCTTTCAG CAGTGACGCA ACATCATTTA AAAGCGGATC TTTATGACAG	9960
25	ATACTTTCGA TGAGCGGTTT CATTCTCATT AACAAATCCA CTTGCTCTTC TCGCATATCA	10020
	AAATAATGAT AGTATGAATT TTCGTTTCTA ACAAATGAT TTTTAACATC TCGGAACGCG	10080
	ATAGACTTCG CCTTTTTAAT ATTTAAAAGT AACACTTCAA ATTCAATCGC AATGGTATCT	10140
30	TCATATTTTT CACAAATATA ACTATATTTA CTAAAAATAT CAGCAATTTG TTGCTCAATT	10200
	TTACATTTGT ATTCGCTCAG TTGTTTGTCT AAACCTGGCA TCATTAAATT CaTTGTAAAT	10260
	GCAATGCTTA GTCCAATTAA CAGTAATAAT GTTTCATTAA CAATTAAATG TGCATCAATT	10320
35	GATTTTGCAT TAAAAACATG AAGTAATATA ACGCAACTCG TAATGACACC TTCTTGTACT	10380
	TTTAATACGA CAGTTAATGG TATAAATAAC AATACGATAA TACCGAGTAC AATTGGACTC	10440
40	TGACCTAATA AACTAAATAT TGCTGAACCT AAAACAATA CTAAAAACA TGATACTAAT	10500
	CTTGAAATAA TCGCTTGTAG CGAATGTACT TTTGTATGTT TAATACATAA TACGACTAAT	10560
	ATGGCGCTTG AAGCATAATT ATCTAAACCT AACAGCTTAC TAATAATTAC ACCTAAAGTC	10620
45	ATACCCACTG CTGTTTTTAT TGTCTAAAT CCAATCTTGT AAGGATTTAA CTTTAACATG	10680
	GGTTAGCGCC TCTTATCTTT CTTACAATA TTTATTGAAT AATGTTTGTA ATTGATTAAT	10740
	TACGTTATC ACATCATGAC CTTGATTTG ATGTCTTTCA ATCATTTCTG TAATCTTTCC	10800
50	ATCTTTTACT AATGCAAATG ACGGACTTGA AGGCGCATAA CCTTCGAAGT ATTCACGCGC	10860
	TCTTTGTGTC GCTTCTTTAT CTTGTCCAGC AAATACTGTC ACTAGACGAT CAGGTAATAC	10920

55

	AGAATTGATC ATAAC TAGTG TTGTACCATC TTGTTTAAGA ACTTTGTCAA CATCTTCTGC	11040
	AGTAGTTAAT TGCTCATATC CCGCAGATT C AATTTCAATC CTTGCTTGTT CTACAACACC	11100
5	GTTTCATGTAT AAATCGAAAT TCATGnCCAT AAGTTCAATC ACCTATCCCT TTATATTTAA	11160
	ACTAtCCTCA TTCTACTAAT TAATAACATA TTGTTCAATA AACTAATCTG AATCACACCT	11220
10	ATATTTAGAC ACAATTTTAA CAATATACCA AACATTATTG TGCTTAAAAT CATGGTAACT	11280
	AATTTGTTCA CATGTTTTCA TTAATATGTT TCAAGTATGA TGTCTTATTT TGACTTTACT	11340
	GCAAAAATGC ATTCAACCAT GTTGATTATT GTTCTTTATC TTTTGTGAAT ATATTGCACA	11400
15	TATTTTAGTG CCAAAAATA ATACATCCAT CGACAAGAAC AAGATAAAAC AAGTTGTCGA	11460
	TAGATGCATC TATGTTATCA CTAATATATA TTTGTATTTT CTAAAGTATA CTGTTGATA	11520
	CGCTGTTTAA TATGATTCAT AATTTTACCT GTTTGTAAAC CATCTAAAAT ACGATGATCA	11580
20	ATTGAAATAC ATAAATTAAC CATGTTACGA ATTGCAATCA TATCATTAAT TACTACTGGC	11640
	TTTTTAACGA TTGATTCTAC TTGTAAAATC GCTGCTTG TG GATGATTTAT AATACCCATT	11700
	GATGATACTG AACCAAATGT ACCAGTATTA TTTACCGTAA ATGTACCGCC CTGCATATCT	11760
25	TCAGCTGTCA ATTGCTTATT ACGCGCTTTC GTTGCTAAAG TATTAATTTT TCTAGCTATA	11820
	CCTTTGATTG ACTTTTCGTC TGCATGCTTA ATCACAGGTA CGTATAATTT ATTTTCATCA	11880
30	GCAACAGCAA TTGAAATATT AATGTCTTTA TGTAAGACAA TTTCAATTTCC TTGCCAGCTA	11940
	CTATTTAATA AAGGATATGC TTTTAAAGCA TCTGCTACAG CTTTACAAA GAAAGCAAAG	12000
	AACGTTAGAT TATATCCTTC TTTATTTTTA AAGCTGTTTT TATAATGATT TCTCGTATTC	12060
35	ACAAGATTTG TAGCATCTAC TTCAATCATC ATCCATGCAT GTGGAATCTC TGTTACACTA	12120
	TTAACCATAT TTTGCGCAAT TGCTTTACGC ACACCATTTA CTGGTATTGT GCTGTTTTCA	12180
	CTATTTGCTT CAGATGATTG GTTACTTGAT GTATCTACTG ATGTTGATTT TGTTTGAACT	12240
40	TGTTTGTG CAG ATTGAGCTGT GGTACCACCA TTTTCAATAA CTGACATTAT ATCCTTCTTA	12300
	GTTACACGAC CTTCAAATCC ACTACCTACA ACTTGTGATA AATCAATGTC ATGCTCTGAA	12360
	GCGAGTTTAA ATACAACAGG TGAAAAGCGA CCATTATTAC GTGGTTGATT TTGTTTAGCA	12420
45	GTAGATGTCT GTTCCACTGT TGCACTAGCT TTTT TAGTAG ATTTCTGAGT ATGCTCATCC	12480
	ACTTTTGCTT GTATCTCTTC AGTTGTTTCA TTTGTCTTTT CATCAGCAGT TTCAATTTTA	12540
	CAGATAATTG TATCAATAGC TACTGTCTGC CCCGCTTCAA CTAAAATTTT TGTAATTGTT	12600
50	CCTGATATCG TGGAAGGGAC TTCAGCTGTC ACTTTATCTG TAATAACTTC ACATAATGGT	12660
	TCATATTCAT CAATATGATC ACCAACAGAA ACTAACCATT GTTCAATGGT GCCTTCATGA	12720

55

	AATTCACGCA TTTTATTTAA GATTTTTTCT GGATTCATCA TAATTTTCATT TTCTAATACA	12840
	GGAGAAAATG GCATAGATGG TACatCTGGA GCAGCTAAAC GCATGATTGG TGCATCTAAA	12900
5	TCGAACAAGC AATGCTCTGC AATAATCGCT GACACTTCTG ACATAATACT ACCTTCTAAA	12960
	TTATCTTCAG TTACAAGTAA AACTTTACCT GTATGTTTAG CACGATCAAT AATTGTTTCT	13020
10	TTATCTAATG GATAAACAGT TCGTAAATCA ACGACTTCAA CATTGATACC GTCTGCAGCT	13080
	AAAATATCCG CTGCTTGTA ACAATAATTG ACCATTAATC CATAACAAAA TACTGTATAA	13140
	TCTTCACCTT CACGTTTCAC ATCTGCTTTT CCTAAAGGTA CAGTGTAATA TTCTTCTGGC	13200
15	ACTTCTTCCT TTAAGAAACG ATAAGCTTTT TTATGCTCAA AGTACAATAC TGGATCATTT	13260
	GATTCGATAG ATGATAATAA AAGCCCTTTA GCATCATACG GTGTGGAAGG AATAACAATT	13320
	GTTAAACCTG GCGATGAAGC AAATATACTT TCAATACTTT GTGAATGATA TAGTCCTCCG	13380
20	TGAACACCGc CACCAAATGG TGCACGAATC GTTAATGGGC ATTGCCAATC ATTATTTGAA	13440
	CGATAACGCA TTTTCGCAGC TTCACTAATA ATTTGATTTG TCGCAGGTAA AATAAAATCT	13500
	GCAAATTGAA TTTCTGCAAT TGGTCTTTTA CCTACCATAG CTGCACCAAT GGCAGTTCCA	13560
25	ACAATATTTG ACTCAGCTAA TGGCGTATCG ATAACCTCTGT CTTCAACATA TTTTGTGTGC	13620
	AGTCCTTGAG TAGTACCAA TACGCCACCT TTTCTACCAA CATCTTCACC AAGAATAAAC	13680
30	ACATCTTTAT TTTGTGTAA TGCTAAGTCT TGTGCCTGcG TATCGCCTCT AAATAAGATA	13740
	ATTTAGCCAT TAGTTAAGAC TCCCTTCTTC GTACACAAAT GCATAGGCTT CTTGACACT	13800
	TGGATATGGC GCGTCTTCAG CAGCCTTTGT CGCTTTATTG ATGATGTCTT TnATgTCCGC	13860
35	TTCTATTTCT GCCAACCAAG CATCATCGAT AATGCCAGCT GAAAGCAACT CTTTTTTGAA	13920
	CTTTTCATTG CAGTCTGCTT TTtAAGcGT TTCACGCTCT TCTTTCGTAC GATATTGGTC	13980
	GTcATCATCT GATGAATGAG CTGTCATACG ACTTGTTACT GCTTCAATCA AAGTTGAACC	14040
40	TTGACCAGAA ATAGCTCGAT CTCTTGCTTC TTTCATCGCT TTATACATTG CTAATGGATC	14100
	ATTACCATCT ACTTGTTTAC CATGTATACC GTAACCAAGT GCTCTATCCG ATAATTTTTC	14160
	AGCTGCGTAT TGTAATGAAT CAGGTACTGA AATTGCATAT TTATTATTTA TAATGACACA	14220
45	TACAAAAGGA AGTTTGTGTA CACCCGCGAA GTTTAAACCT TCATGGAAGT CACCTTGTTT	14280
	TGAGCTACCT TCACCAACAG TGCTGTTGC AATTTTCTTC TTACCATCCA TTTTAAAGC	14340
	TAAAGCAGCA CCAACAGCAT GGGGTATTTG AGTTGCTACC GGTGAACTTT GAGACAAAAT	14400
50	ATTCTTAGCT CTACTIONTAA AGTGTGATGG CATTTGTTTT CCACCAGAGT TAACATCGTC	14460
	TTTCTTTCCA AACGCTGATA AAAACGTATC ATACGCTGAG ATACCCATAT AAGTAACGAA	14520

55

AATCTGAGTT GCTTCTTGTC CTTGACCACT TACAACAAAT GGAATTTTAC CTGCACGGTT 14640
 CAATAACCAC AGTCTTTCAT CTATTTTCT ACCTAAATCC ATCCATTAT ATATTACTTT 14700
 5 TAGGTCTTCT TCGCTAAGGC CTAATGATTT ATAATCAATC ATGTTAAATC CTCCTATTTA 14760
 TACGTGAATA GCTCTACTTT CTGCTTTCAA TCCTAATTCC ATCAACACTT CAGAGATGGA 14820
 10 AGGATGTGCG TGTGTTGTTA GTCCTAATTC TAATGCCGAG CCATTTCATGA ACTGTAACAG 14880
 TGATGCCTCA TTAATCAATT CTGTTACATG TGGACCAATC ATATTAATAC CCACAATTTT 14940
 TTCAGTTGAT TGATCAATCA CCATTTTCGCT ATACCCTTCG TTTGTGTCAT GGCTATCAAT 15000
 15 CACTGCTTTA CCAATTGCTT TAAATGGTAC TTTAAAACCTT TTAACCTTCA TTCCCTCTGC 15060
 CTTTGCTTGT TCAATGTTTA AACCGATAGA AGCAATTTCA GGTGTGTAAT AAATACACTT 15120
 AGGCATCATG TTATAGTTTA CTGGGATTGG GTTCCCCTCA AACATATGAT CAACAGCCAC 15180
 20 AACACCTTCT TTTGATCCAA CATGTGCCAA TTGTAATTTT CCTATACAAT CACCAGCTGC 15240
 ATAAATATGT TTATCTTCAG TTTGTTGAAA TTCGTTTCGTT AAAATATGTC CTGATGTTGa 15300
 AAGTTTATT TTAGTGTTGT TTAACCAAT ATCTGATGTG TTAGGTTTTC TACCAATCGA 15360
 25 TAGCAACACT TTATCTACTT TAATTATGTC TGAGGAAAT TCAAACGTAA CACCATCTTC 15420
 GTTAACATTT ATATCATTTT CAGAAAGTTT TATTCCCTCA TAGAATTTAA CACCACGTGC 15480
 TGACAATGAT TTTTAAATA GTTGTGAAGC TTGTTTACTT TCAGTTGGTA AAATTCTTTC 15540
 30 ACCTGCTTCT ATAAGTTTA CGTCAACACC TAAATCTATC ATCAATGATG CAAATTCCAT 15600
 TCCGATAACA CCACCACCAA TAATACCAAT ACTTGATGGT AACGTCTTCA ATGATAATAT 15660
 ATCATCGCTA GATAAAATTT TATCATGATC AAATGATAAG AATGGCAACT CTGCAGGCGA 15720
 35 AGAACCAGTT GCAATTAATA CAAATTGGTT GGGTAATAAG TCTGATTCAC CATCTTCATA 15780
 TTCGACAGAA ATTGTGCCAC TTTGAGGTGA AAATATAGAT GTACCTAGAA TACGTCCCGT 15840
 GCCATTATAA ATGTCAATGT GATTGTGTTG CATTAAATGC TTTACACCTT GATACATTTG 15900
 40 ATTAATAATG TCTTCTTTTC GTGCCAACAT ATTTTCAAAA TTAACATTAG CATCTTTGAC 15960
 ATCAACGCCA AACATTGCTG CCTGTTTTAC TGTTTGAAAT ACTTCAGCAG ATTTAAGCAG 16020
 45 CGATTTAGTA GGAATACAAC CTTTATGGAG ACAAGTACCT CCTAATAGTT GTCGTTCTAC 16080
 TATTGCCACT TTTTACCTA ATTGAGACGC ACGTATCGCA GCAACATATC CTGCAGTACC 16140
 TCCACCGAGA ACGACTAAAT CATATTGTTT CTCTGACATG TTCTTACTCC TAACTAATGA 16200
 50 TATATATCCA TTGAAAATTT ATTAATACAT AGTTTTCATG TCCATTAAAT ACCTATTTTA 16260
 CATGATTGTC TATTTAGTTT GAATGCACAT AAATAAATCC ATAAATGAGT ATTCAACACA 16320

55

TAAATCAGTA ACACTTGCAC CTGAAATCAT TCGTGCAATT TCATCTACTT TATCATCGCT 16440
 AATTAECTCT TGAACCTGTG TTGTTGTACG ATCATCTTTT GATGATTTCG AAATTAATAA 16500
 5 ATGATGGTCG CTCATCGATG CAACTTGTGG TAACTGAGAG ATACAAATAA CTTGTATATA 16560
 TTCTGCTaTA TCTCGCATTT TCTCTGCCAT TT 16592

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

20 CCAATACAAC GTAAAAAGAT TGCTTGTGTT ATTAATGAGT TAGATAAAAT AATTAAAGGA 60
 TTTAATAAGG AAAGAGACTA CATAAAATAT CAATGGGCTC CAAAATATAG CAAAGAnTTT 120
 TTTATACTTT TTATGAACAT TATGTACTCA AAAGATTTTT TAAAATATCG ATTTAATTTA 180
 25 ACATTTCCTG ATTTATCTAT CTTATATGTA ATATCATCTC GAAAAAATGA GATACTAAAT 240
 TTAAAGATT TGTTTGAAAG TATTAGATTT ATGTATCCTC AAATTGTTAG GTCAGTTAAT 300
 AGATTAAATA ATAAAGGTAT GCTAATCAAA GAACGATCCC TTGCAGATGA AAGGATTGTG 360
 30 TTAATCAAAA TAAATAAAAT ACAATATAAC ACTATTAAAA GCATATTACAC AGATACTTCC 420
 AAGATTCTCA AACCAAGAAA ATTTTTCTTT TAAATTTAAA CAGATTTACC TCTTGATAAA 480
 35 ATAAATAAGC AATCATACTA CTTCTCAATT TAGTATAAAT AAAAATACAT AATTAECTTT 540
 CTTTTGTTTT TATATTATTT CAATACCCTA CTATATATCA CAACACATAA ATTAAGCATG 600
 ACACTCATTG AATTTAGTTC ACCATTTCTG GTTCCAATTT TACTGAGTAT CATGCTTTTA 660
 40 ATGTTATAAA CCTAATGCTT TAATAAATCG TGTTAATTCT TCTCGCATAC TGTCATCTTT 720
 CAATGCATAT TCTATGGTAG TTTTAACGAA GCCTAATTTT TCTCCAACGT CATAACGTTT 780
 GCCTTCGAAG TCATATGCAT ACACTTGGTT ATCATTATTC ATACGTTCAA TCGCATCTGT 840
 45 TAACTGAATT TCGTTACCTG CGCCTTCTTT TTGCGTTTTT AAATAATCGA AAATTTTCAGG 900
 CGTTAATACA TAACGTCCCA TAATAGCTAG GTTTGATGGT GCCGTACCTT GTGCTGGCTT 960
 TTCAACAAAC TTTTCACTT CATACTGACG TCCGTTTTTA GTTAATGGGT CAATAATTCC 1020
 50 ATAACGATGA GTATCTGCTT CCGGAACCTT TTGGACACCT ATAACGAGT GCCCTGTTTC 1080
 TTCATAAACG TCAATCAACT GTTCACTGCT TGGCACTTCA GATTCAACAA TATCGTCACC 1140

	TAAACCTTTT	TGTTCTTCT	GCCTTACATA	AAAAATATTC	GCAAGTTCCG	TTGAATACTG	1260
	AACCTTCTCT	AGTAATTCAG	ATTTACCTTT	TTCTTTTAAC	ACCATTTCTA	ATTCTTTTGT	1320
5	ACTATCAAAA	TGATCTTCAA	TCGCGCGTTT	GTGGCGACCT	GTCACATATA	TAATATCTTC	1380
	AATTCCAGCT	CTTGAGCTT	CTTCAACGAT	ATATTGTATT	GTGGGTTTAT	CTAAGATAGG	1440
	AAGCATTTCC	TTTGGCATCG	CTTTAGTTGC	TGGTAAAAAT	CTAGTCCCTA	AACCAGCAGC	1500
10	GGGAATGATT	GCCTTTTTTA	TTTTTTTCAA	AGTTAATGTG	CTCCTTTTCC	TAAGTATTAA	1560
	ATCTATGTAT	CAACGTCATT	TTAACACTAA	TTAGAACGCC	TTCATAGTGT	CATTGAGTAT	1620
	GTAATTATTT	CTTGGGAAAT	TTGTTTTAAT	TTTAAAAAAC	AGGCTTACTT	CATATAATTT	1680
15	ATGAAATAAA	CCTGTCAATT	TTGGATTGAT	TATGCTTTGT	GATTCTTTTT	ATTTCTGCGT	1740
	AATAACGCTA	AACCTAAAAAT	GCTAAATAAT	CCGCCGAACA	ACATGCCGTT	GTTTGTGAT	1800
20	TCTTCTCCAC	CTGTTTCAGG	TAGTTCAGAT	TTCTTAGATT	GTGCTTTTTT	AGTTGGTACC	1860
	ACTGCTTTAA	CCTTTTCATT	GATTTCAATA	ACAGGTGTTA	CTACTTTACC	TTGTTCCACT	1920
	GGTTTAGAAG	GTTTTTTAGG	TTCTTCTTTA	GCAGGTGGTA	TTGGTTTACC	AGGTTCAGTT	1980
25	GGTACCTCTG	GCGTTGGCGG	TGTTGGTGTT	TCCGGCTCGC	TTGGTACTTC	TGGTGTCCGT	2040
	GGTGTGGTG	TTTCCGGCTC	GCTTGGTACT	TCTGGTGTGC	GTGGCGTTGG	TGGCACGATT	2100
	GGAGGTGTTG	TATCTTCTTC	AATCGTTTGT	TGACCTTCAT	TATGACCACT	TACTTGTGGA	2160
30	AGTGTATCTT	CTTCAAAGTC	AACACTATTG	TGTCCACCGA	ATTGATAATT	TGGTTTATCT	2220
	TTATTTGTAT	CTTCTTCAAT	AATTTCAAGT	TGCTTATTGA	ATCCGTGAAT	ATGTGGCACA	2280
	CTGTGGAAGT	CGATATCAAT	GATATTACCA	CCTTGTTTCAT	ACTTAGGTTT	GTCTTTCTCT	2340
35	GTATCTTCTT	CGAATGATTG	GTTACCATTG	TTTGGACCAT	GAATTTGAGG	TACACTATCG	2400
	AAATCGATAT	CTACGATATT	GCCACCTTGT	TCATATTTTCG	GTTTATCTTC	TTCTGTGTCT	2460
40	TCCTCAAATG	ACTGATTACC	GCTATTTTGG	CCACCTTCGT	AACCTAATTC	ACTCTTAATA	2520
	TCCACGTGGC	TATTTTCTTC	GATTTCTTCA	ATCACGCCAT	AATTACCGTG	ACCATTTTCA	2580
	GTTCTTAAAC	CAGAATGAGA	AATATGATGA	TTGTTTTTCAG	TAATTTCTTC	GATTGGTCCT	2640
45	TGCGCTTGAC	CATGTTCTTC	AGGTAGTTCA	TCTACTAGTT	CAATCAGATT	ACTTTTCAGTC	2700
	GTATATTCTT	TCGTATCTTC	AATTGTTGTA	TGATCGCTAA	CAGCACCAGT	TACAATACCT	2760
	TTTGTAGAAT	CTTCGTCAAA	TTCAACTAGG	TTAGACTCAG	TAGTAACCTG	ACCACCACCT	2820
50	GGGTTTGTAT	CTTCTTCATA	TTCAACAACA	TCAGCATGAT	GTTTTGAATT	TTCATGTGTC	2880
	GATTCTTCAA	AGTCTACATG	AATAGAATCT	TCTTCAGTTT	CAATGGTACC	TTCTGCATGA	2940
55							

	TCTTCGATTG TACCAAGTCAA TTCATGCTTC TCCACTGGCG GCTCTGATT T AAATTCAAGT	3060
5	TCGATAGGAG TACTATGTTC TATAATAGGT TCCTTTAGTT TATCTTTGCC GTCGCCTGA	3120
	GCGTTATTAG AGTAAAATGC AACGCCATTT TTCCaAGTTA AATTACTTGT ATAATAATAG	3180
	TTATAATATC CAAAAAGGTG TGTTTGAAAT TCTAAGTTGC TAGCATTGA ATCATAATAC	3240
10	CCTTCATATT TTATTACATA ATTTTACTT TGGTCTAAAT TATTAAAGTT TAAAGAATAA	3300
	CCACCATTAG TATCAAAATC TAAACTCATA TTATCAGTCA CATCTTCAAA TTTGCTGACA	3360
	TCATCAAGCT TTGCATAnTn AgctTTCAGC TAAATCGTCT GAACCAATGT GTTTATATAC	3420
15	CTTAACTGTT GGATTATTAA CCCCTGGTTT ATTTCTTTA GTTACTTGAC CAGTTACTGT	3480
	CACAGAGCTT AACGACTGGT TGTTAGGTTT CATGTACGCA AAATGACTAA ATTTCCCATC	3540
	TACTTTATTT AAAGTATCAA TTCGACCATT AGCTGTTACT CCCCAATTAT CTCTAACTCC	3600
20	ACCTAAATAT TGAATATTAA ATATTTTGCT AACCGTAGTC TCACCCAATT TAACCTCAAC	3660
	ATTTTGGTTA CCTTTTTCG TCACTGTTGT AGGATCAATA AATAGATTTA AAGATAATTC	3720
	AGCAGTTAAA TCTTCTTTT CTGTACATA TTCTTTAAAC GTATATCTAA CTTTCTTTC	3780
25	TCCAATTATT TCTCCTGTCG CCATAACTTG ACCATCTGTA CTTTTTATCT CCGGAACTTT	3840
	ACGCAGTGTT GAGATACCAT GAGTTTCAAC ATTATCGCTT AATGTGAAAT CAAAATAATC	3900
30	TCCCGCCTTA ATTCCTTCTC CAAATTTCCA TTATATTTT AAGGTTACTC TTTCTGCGTT	3960
	ATGAGGATTT ACAACATTGC TATCTTGTTT ATGTCCTACA ATTTCACTAC CTTCTTCTAC	4020
	TTCCACTTTA TTTGTTACAT CTGTACCTGT CGCTTTAGTT TCTTCCACTA CTTCTTCTC	4080
35	TGCAACTGCT GTAACGTCA TgatCTTTTC ATTCTTGGTT TAATTTCTGA GACGTTACTT	4140
	GGTTGAGCTA TGTCAACTTG AGTTCCTGTA GTTTCCTTAT CAGCAACTTT TTCCGATGGC	4200
	AAATCAACTC GCGAAgTTTC TACTTTTGGT GCTTGCAcAG TTTTCGGTGC TTCTTCTGTT	4260
40	GTTACTTG TGATTGTGA TGGTTGCTCA GTTGATGTCG CGCTGTATGA TTGTGTTTCA	4320
	TCTATGTAT TAACGTTATT TGTAGTTGTT TGTGTTTCGC TTGCTTTACT TTCAGTAGCT	4380
	GAACCTCCAC TTTCTCTAC TGTAGTATTG TTTGTTCCG ATGCTGCAGC TTCTTTTCT	4440
45	TGTCCCATTC CAACAACGAT CATGTTTCCT AAGAATACTG AGGCCGCTCC CAATTTGTGT	4500
	TTTCTTATGC CGTATCTAAG ATTGCTTTTC ACTATAATAT TCTCCCTTAA ATGCAAAATT	4560
50	CATTTATTTT TAAAAC TCAA TAAATGCAAT TCTATATTGT TCGGTTTTTA AAAGCAATGA	4620
	AAAAAAGCGA GTTAATAAAA AGTTAAGATT GTTGTTAACT TTATGTATAA TGAGTTTTTT	4680
	ATTATTTGAA ACTCACATAT ATATTGCATA CAAAGCTCTT GAACACCTTG ATATAACAGG	4740

55

	TACTAAACCA TACATAATAA TCGCCTGTAC AATGCATCAT TAACAAGTCA CTGAAACGCC	4860
	TTTCATTGTA TTAATAACGT CACTATAATT TTTATATCGT TCGGTTTTTG TTTGATTTTA	4920
5	ATGATTATTT ATACAAAAAC AGCCGTATTT CAAGCCGACA TTTTAAATTT AACTAAATTT	4980
	GCATCTAGTT AATAATTGCA TTTATCAAAT TTGTCTTATT GATCCAATCT AATTTGTACT	5040
10	CACAACTAG TTTAAATTC TAACCTTATC TCTCAGTTCG TTATCAATCA TCAGACATAA	5100
	ACCAATGAAG CAATCAGAAA ACACTCTAAT TTTCTATTAG AAATTGATT TAATATAAAA	5160
	AAACAGGCTT ACTTCATATA ATTTATGAAA TAAACCCGTC AATTTTGTGTT TAATTATGCT	5220
15	TTGTGATTCT TTTTATTTCT GCGTAATAAT GCTAAACCTA GAATGCTGAA TAATCCGCCG	5280
	AACAACATAC CTTTGTGTTG TGATTCTTCT CCACCTGTTT CAGGTAGTTC AGATTTCTTA	5340
	GATTGTGGTT TTTTAGTTGG TGCCACTGCT TTAACCTTTT CATTGATTTT AATAACAGGT	5400
20	GTTACTACTT TACCTTGTTT CACTGGTTTA GAAGGCTTTT TAGGTTCTTC TTTGGCAGGT	5460
	GGTACTGGTT TACCAGGTTT AGCTGGTACC TCTGGTGTG GCGGTGTTGG AGTTTCTGGC	5520
	TCACTCGGCA CTTCTGGTGT CGGTGGTGTG GGTGTTTCCG GCTCACTTGG TACTTCTGGT	5580
25	GTTGGTGGCG TTGGTGTTC CGGCTCACTT GGTACTTCTG GTGTCGGTGG CGTTGGTGGC	5640
	ACGATTGGAG GTGTTGTATC TTCTTCAATC GTTTGTGAC CTTCATTTTG GCCGCTTACT	5700
	TTTGGAAGTG TATCTTCTTC AAAGTCAACA CTATTGTGTC CACCGAATTG ATAACCTGGT	5760
30	TTATCTTTAT TTGTATCTTC TTCAATAATT TCAGTGTGCT TATTGAATCC GTGAATATGT	5820
	GGCACACTGT CGAAGTCGAT ATCAATGATG TTACCGCCAT GTTCATACTT AGGTTTGTCT	5880
35	TTTTCTGTAT CTTCTCGAA TGACTGATTA CCTTTATTTT GACCATGAAT TTGAGGTACA	5940
	CTATCAAAAT CGATATCTAC GATATTGCCA CCTTGTTTCAT ATTTAGGTTT GTCTTCTTCT	6000
	GTGTCTTCT CGAATGACTG GTTACCGCTA TTTTGGCCAC CTTCATAACC TAATTCACTC	6060
40	TTAATATCAA CGTGGCTATT TTCTTCGATT TCTTCAATCA CGTCATAATT CCCGTGACCA	6120
	TTTTCAGTTC CTAAACCAGA ATGAGAAATA TGATGATTGT TTTTAGTAAT TTCCTCGACT	6180
	GGTCCTTGTG CTTGACCATG CTCTTCAGGT AATTCATCCA CTAATTCAAT CAGATTACTT	6240
45	TCAGTTGTAT ATTCTTTCGT ATCTTCAACT GTTGATGAT CGCTCACTGC GCCAGTTACA	6300
	ATACCTTTTG TAGACTCTTC GTCAAATTCA ACTAAGTTAG ACTCAGTAGT AACCTGACCA	6360
	CCACCTGGGT TTGTATCTTC TTCATATTCA ACAACATCAG CGTGATGTTT TGAATTTTCA	6420
50	TGTGTAGATT CTTCAAAGTC AATTGGATTT GATTCCTCAG AGGACTCAGT GTATCCTCCA	6480
	ACGTGACCTG cTTCGCTATC CACAGCAGTA TGGTAATCGA TATCAATAGC TGATGAATCC	6540

55

	TGGTAATCAA	TGTCAAGAGT	TGATGAATCA	TATTCCTCTT	CAACAGTAGT	TACTAAATTC	6660
	TTATCATATT	GACCTGTAAG	AGTTTCTTTA	ATTGTATCTT	CTTTATATTC	AAATTTATTA	6720
5	TTTTGAATAA	TCGGACCATT	TTTCTCATTT	CCGTTGCTT	TATTACTGTA	TAAAACTAAA	6780
	CCATTATCCC	AAGTTAAGGT	ATATCCTCTA	TCATAATAAT	ACTTATAAAG	TTGCTCTGGA	6840
10	TGTCCTACCA	TTGTGTTCT	AAAATCAACT	TCATCAGTAC	CATTTAAATA	CTCTCCATCA	6900
	TAGTGAACAA	CATAAGTTTT	ATCTAGATTT	TCTATATTCA	ATGAATAGCT	TCCATTATTT	6960
	TGTAAATTCA	AATTCCTACT	CATATTACTT	GTGACTTCTT	TAAATTTAGA	AGTATCTGTC	7020
15	GTATTTGCAT	ATACACTCTT	CGCTATGTCT	TCATTATTAC	CCAAGTATTC	AAATATCCTA	7080
	ACTTTTGGTT	GATTTCCATT	CTGATTACTA	CCTTTTCATTA	AAGTTCCAGT	AACAGTCACA	7140
	CTTGTCGTTT	TACCATTATT	AGGTTTAATA	AATGCAACAT	GCGAAAATCT	ATTATTCGCT	7200
20	TTATTAAATG	TCTCAATCGA	TCCATTTAAA	TTGGCATAAT	AATTCCTCAAT	ACCATCTTTA	7260
	TATTTAACAT	CTAATTCCTT	TGAAGTTTGT	TCTTCATTTA	GTGTTGAAGT	TATAGTTTGA	7320
	TTTCCATTAG	TTGTACAGT	TTAGGATCA	ATAAATAAAT	TAATTTCTAG	TTCAGCCGTT	7380
25	ACATCAACCT	TATCTTCAAT	ATCATTTGTA	AATGTATATC	TAATCTTTCC	ACCTTCTAAA	7440
	ACTTCACCTG	TCGCCATTAC	GACTIONACCA	TTTTTAATTT	CTGGTACTTT	TCTAGCAGTT	7500
	GATACGCCAT	GCGTATTTAC	ATTATTTGAT	AAAGTAAAGT	CAAAGTAGTC	ACCTTGATGT	7560
30	AAACCATTCT	CAAATTTCAA	CTTATATTTT	AGTACCGCTC	GTTGTCCTGC	ATGAGGTTCT	7620
	ACTTTATTTG	TATTGTTATG	CCCCTCAATA	GAACCAATTT	CTACTGTAAC	TTTACTTGTT	7680
35	ACATCTGTAC	CCGTTTCCAC	TTTCGCGTTA	CTAGCTTCCT	TAGCTTCCGC	TACATCTGCT	7740
	GATCTTGTC	CACGTGGCTT	ACTTTCTGAT	GCCGTTCTTG	GCTGTGCCAC	TTCAACTTGT	7800
	GTTTCTGCGA	CTTGATTTTG	TGTAGCCTTT	TTAGGTGTTA	AATCTIACCTG	TCTTTGATCT	7860
40	CCGCTATTGT	CTTGAGATTG	TGTTGTTTCC	TTAACTTGAG	GTTTCGCTTC	TTCTTAACT	7920
	ACCTCTTCTT	TAACTGTTTC	TATATTTGCT	GTTTGTGCAG	TTTGTGGTGC	TTGTACTGCT	7980
	TTTGGTGCTT	CTTCAGTTGT	TACTTGTGTT	GCGTTTGACG	GTTGTTCTGT	TACTGTTGCG	8040
45	TTATATGATT	GAGTTTCTTC	TATATGATTA	ACGTTAGTTG	CAGTTGTTTG	TGTTTCACTT	8100
	GTTTTATTAT	CAGTAGCTGA	ATTCCCATTT	TCTTCTACTG	TAGTTGTCTT	TTGTTCTGAT	8160
	GCTGCAGCTT	CTTTGTCTTG	TCCCATCCCA	ACAACGATCA	TTGTTCTCTA	GAATACTGAT	8220
50	GCTGCTCCCA	ATTTATGTTT	TCTAATGCCG	TACCTAAGAT	TGTTTTTCAC	TATAATATCT	8280
	CCCTTTAAAT	GCAAAATTC	TTAATTTTTT	AAACTTAATA	AATGCAAGTC	TATATTGTTT	8340

55

	ATGTTAATTG ATAATTTTAT TATTTGAAAT ATACCTATAA ATTGTATTCA AGTCATCAGA	8460
	AACCCTTGTC ACACAAGGCT TGTATTTTTT ATACTTATTT TTAAATTAA ATTCATCATT	8520
5	ATCTAATTTA AAACAATATA CTAAACGTTT CATAATTATC GCCTGTACAA TACGCACAAA	8580
	AACATGTCTT GAAACGCCTT TCATTACTCT AAAATACCCA ATATACTTTT TATATCGTTC	8640
	GGATTCTGAG TATTTTCAGAC GATTTTCTGC ATAAAAATAA ACGTGTTTCA AGGCAATATA	8700
10	TTGCAATTAC CTAAAAACAC GTTTACTTAA TATTIAGTTA AACAAATAAG CTAATGAATA	8760
	AAATGAAGAT GATACCTGAA ACGGAAATAA TCGTTTCTAA TAATGACCAT GTTAAGAATG	8820
	TTTCTTTTAC AGTTAAACCA AAATATTCTT TAAACATCCA AAATCCTGCG TCATTTACAT	8880
15	GAGACAAAAT CACACTACCT GCACCTATCG CAAGTACAAC TAATGCAACA TTTACATCTG	8940
	ATGATTGTAA TAATGGTAAG ACAATACCTG TAGTTGAAAT CGCAGCTACT GTAGCCGAAC	9000
20	CTAATGCGAT ACGTAGCACA GCTGCAACAA TCCATGCTAG TAAATCGGA GACATCTCTG	9060
	TACCTTCAAA CATTTTAGCA ATTGTATTTT CGACACCGCC GTCAATTAAT ACTTGTTTAA	9120
	ATGTACCGCC ACCGCCAATA ATCAATAACA TCATTCCGAT TGGATAAATC GCATTTCGTCA	9180
25	CTGATTCCAT AATATGATTG ATCTTACGCT TTCTCATTAA TCCCATCGTA ACGATTGCAA	9240
	ATAATACTGC TATTAGCATG GCTGTCCCTG CTGTTCTTAT CATATAAATG ATAGATTCAA	9300
	ATAGATTGTG AGGTTTGTCA TGCCCAGTTA CAAGTTGCGT TATCGTAGAC ACTAACATTA	9360
30	ATATGACTGG TAATGTTGCT GTTAATAAAC TCATACCAAA TCCTGGCATC TCTTGATCCG	9420
	TAAATCTTTT TTGTGCACCT AACGCTGAAA TATCGCCTTC TCGTGATAC GCAGACGGAA	9480
35	TCATTTTTTG TGCACTTTGT TAAATATAGG CCCTGCAATG AGTGTAATG GaATGGCAAT	9540
	AATCATACCA TACAGTAATA CATCTCCAAC ATTTGCCTTT AATTCTTTTG CGATGACTAC	9600
	CGGTCCTGGA TGTGGTGGTA AAAAGCCATG TGTCATGAT AAAGCTGTTA CCATAGGTAG	9660
40	TCCTAGTTTT AACACTGAAA CATTTGCGCG TTTTGCTACT GTAAATACTA ATGGAATCAG	9720
	TAAGACTAAA CCTACTTCAA AGAACAATGC AATACCGACG ATAAATGCTG CAACAAGCAT	9780
	TGCCCATTGT ACATGTTTTT GACCAAATTT TTGAATCAAC GTGTCTGCGA TTCGAGTTGC	9840
45	ACCACCACCA TCAGCAAGCA ATTTCCCAAG TATGGCACCT AAACCGAATA TCAGTGCAAT	9900
	GTGGCCGAGC GTACTGCCCA TTCCTTTCTC AATCGTCTCC ATAATTTTAG TCAATGGTAT	9960
	ACCTAGCATT AACGCTGTAA TCATCGATGT GATAATTAAT GAAATAAATG TATTTAATTT	10020
50	AAACCCAATA ATTAATACTA ATAAAATAAC GATACCTAAA ACAACACTGA TTAACGGCCA	10080
	TATTTGTTA AACATGACAT TCCCCTCTTT CTCTTTTCAA TAGAATGTAA CACCGTCGTC	10140

55

	GAGTGACGTA TTTATTGTGT TTTATTTTCA GCGATATGTT GCGTTGAAA ATCTGCAATT	10260
5	TGTTCATAAT TCTCTGTAA AGAACGACTT AAATTGATAA AAATGGATAC GATCTCTTGG	10320
	TAAACAGTGA CATTTTCTTC AATCGGCGTA TGATTGTTTG TGGCACCGAC CATCGATGAA	10380
	ACGATTGAAA AATCTTCAAT GTCACCTACA GCTTTAAGTC CGAGCACGCA GGCACCTAAG	10440
10	CATGAACTTT CATAACTTTC AGGAACCACT AACTCTGTGT CAAATATATC TGACATCATT	10500
	TGACGCCATA CTTCACTTTT CGCAAAACCA CCTGTTGCTT TTATCATCTT AGGTGTTTCA	10560
	TTCATTACTT CAATAAGCGC AAGATAGACG GTATACAAAT TGTAAAGAAC ACCTTCTAAT	10620
15	GCAGCGCGAA TCATATGTTT TTTTTTATGA GATAAAGTTA AACCGAAGAA TGAACCTCTT	10680
	GCAATTTGCGT TCCAAAGCGG CGCACGTTCT CCTGCTAAAT AGGGATGGAA TATTAAACCA	10740
	TCTGCACCTG GTTTAACACG CTTTGCAATT TGAGTTAAGA CATCATAAGG ATCAACACCG	10800
20	AGACGTTTCG CAGTTTCGAC TTCACTCGCT AGCAACTCGT CGCGCAACCA TCTCAATACG	10860
	ACACCACCAT TATTTACAGG ACCTCCGATG ACGTAGTGGT CCTCTGTAA GACATAACAA	10920
	AATATTCTAC CTTTGTAATC AGTACGCGGT TTATCTATCA CAGTACGAAT CGCCCAGAT	10980
25	GTACCGATTG TGACAGCAAC TTCTCCTTTA CCAACACTAT TGACACCTAA ATTAGAAAGG	11040
	ACCCCATCAC TCGACCAAT AACAAACGGT GTATCTTTAT TAAGCCCCAT TAATGTTGCA	11100
30	TAACGTTCTT TCATACCTTT CATCACATAC GTTGTGGAA CTAATTCCGG CAACATTTCC	11160
	TTGGAAATAC CCAGCAGTTC TAATGCCTCA ACATCCCAAT CTAATGTTTC TAAATTAAAC	11220
	ATCCCTGTTG CGGAAGCCAT TGAATAATCA ATGATATATG TATCAAATAA ATGATAGAAA	11280
35	ATGTATGTTT TAATATCTGC AAACCTAGCA GTACGTTGAA ATACATCTTG CCATTCATGT	11340
	TTCATCCAAA AAATCTTCGC TAATGGCGAC ATAGGATGAA TCGGTGTGCC TGTTCGCTGG	11400
	TAAATCGCAT TGCCATCATG CACTTCATTT ATTACTGTTG CATATTTTGC AGCGCGGTTA	11460
40	TCTGCCCAAG TAATATTATT TGTTAATCTT TGATGTTGCT GATCCATCGC AATCAAGCTA	11520
	TGCATTTGCG CACTAAATGA CACAACTTA ATGTCGTCTT TATTAACTTT GGATTCTCTC	11580
	ATAACATATT TAATAGTCAT TAGTACTGCA TCAAATAATT CATCTGGGTT TTCTTCTGAG	11640
45	ACATCAACGT TTGGTGTGTG TAAATCATAG CCTATTGAT GTTTCATGAT AAAAGTTCCA	11700
	TTTTCATCAT ATAAGACTGA CTTGGTACTC GTCGTTCCAA TGTGACACC AATCATATAT	11760
	TTCATGATAA ATCCTTCTTT CTTTCATTTT AATCAACCA AAATCCTTCA ATATCTTTAC	11820
50	CAACATCGTC GAAATTTAAA TGAAACGCTT CTTTCAAAAT TTGACTGTGC TATTGTTCCA	11880
	CTGCATCAAT AAACACTTGA TGATTATGAT GTATGCGTTC AAAATCTTGC GGGTTCTGTT	11940

55

	AAAATGAGTT TAAATATTGA TGATTAGATG CTTTGATTAA TGTTTCATGA AATTCAAAGT	12060
	CATGCTTCGT AAATGATTCT GCATCCTCAA ATTTTACTGC CACTTTCATC ATTTCAAGTT	12120
5	GTTTCTTCAT TTCTTTTACG ATAGGTAGTC GCTCTTGATT TTAACTCTT GAAAATGCAA	12180
	ATGACTCTAA CATCAGTCGC AAATCATACA TTTCTTTCTT TTCTTGTTCC CCAAACGGCA	12240
10	ACACATGTGC ACCCATTCTT TCTAATTGGA TGAGTTGATT TTGTTGCAAT AATTTAAATG	12300
	CATCTCGAAT TGGCGAACGA CTCACATTAA ATTGCTTTGC CATTTGATTT TCAGTGAGTA	12360
	ACGTACCTTC AGCTATGTGA CCATTCACAA TGCCTAAGCG TAATTCTGCC GCGATACCTT	12420
15	CTCCAGTTGT CATACCTTCC AACCATTCTT CTGGATATCC ATACATCATC AAAGTCACTC	12480
	CTTCATTACA CGACATACTT GTATACAAGT ATGTTAATAT AGTTATTATG AGTTTGCAAG	12540
	CGCTTTCTTT ACGAGCACTA AAATAGTGAC CACCCCTTTT CGATTTAAAT TTAAAGGAAA	12600
20	TGGTCACTAT CACACGAATG ATTTAATTGT TATGTTGTAT GTGGGATATT TCTAATTGTT	12660
	CTGTACTCAT ATGCGCTTTA GGTACTTCAA TGCAATAATG CGTTTCATGA CAGTTTGGAC	12720
	ATTCGAATCG ACGTGTGTGTC GCTGTATGTT TCGCTTTGAT AACTGCCAC AAAGATGGTG	12780
25	AGAATATATG CTGGCAGTTA GGACATAAAT AGGCAACCTT TTGTTGGTAA TAAAAAGTAA	12840
	CACCAATGCC ATAACCAATC ATAAATGGTA AAGCAATTAA AAACGGCCAT TTATTTTCA	12900
	TCAAAATTGC ACTTATAATG CTAGAATATT GAATTATTCC TATAATACCA GCACTAATCC	12960
30	AAATGTTACG ACGAATACTT TTCATTTTCA CTGATTTACT CATGACATGC TCTATGTCTT	13020
	TTAAGTGTGT GATTGGAGAC GTCGACGCTT CATTTACGTA ATATTGAACA TTTTAAATTT	13080
35	TGTTTAAATC CGCTTGTTGC TGTTTAACTT GTTGGTTAAT TTCTTGTTGT TTCATAGTTA	13140
	GTAAAGTATT GAGCGTCTTC AAAGTACCTT CACCTTTTAG CAACATATCT ATATCGCTTA	13200
	ACGCACAACC TAAATCTTTA AGCAATAAGA TTAACCTCTA TGTTTGTCGC TGTTGTTCTG	13260
40	TATACACACG ACGCTTTCCT TCTGTAAATC CTTGTGGTTT CAAAATACCT TTGCGATCAT	13320
	AATATTGAAT CGTTCGTGTT GTCACATTGC ATAATTTTGC GAGTTCTCCA GTCGAATAGT	13380
	TAGACATAGA TTCCACCTCC TATAATTACC ATAGTTGATG ACCCGACGTC ACGAGCAAGT	13440
45	ACAATTTCCA CATTTTAAAG AAATTTATTA TACTAGGCGT CTTATTTTTA TGATTTGTA	13500
	CCATGTTGAT TTACAACTC ACTCAAATA AGTAACACAC CTACTAAACA TCTACTCTGT	13560
	TATTTAGAA TGAATTTGTT GTAATTTATC TTCAACTTCA GTAATCTCTG TCGCACATTC	13620
50	TTTCAGTAAA TCTCGATACT TTTCCGTCTC TGCAATGTTT TTATAACGTA TTTTATGTTT	13680
	TAAACTTGcC CACATATCCA TACCTATCGT TCTAATTTGA ATTTCAACAG GCAATACCTC	13740

55

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1059 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```

GGATAAGTTC AGGTAAATTC ATTTCTTTTT CAATTTTGAT TTTCAATTGTT TCCGCCCTTT      60
TAAAATAAAG TTAGTTGCTT CTGTCCTCA TATTCCAAAT CACTTTGCTT TATATATGTT      120
TCAAGCTCTT CCGCTGTATC AAATGTCTTT TTCACACCTT GCCAACCTGG CACGATATGA      180
CCGTGAAAGT AATAAGTGCC ATTTACTACA TGGATATGTG CCACTCGTTC GTTATCCTGA      240
TACAGATATC TCTTAGATCC AAAGAATTGA TTTAGGTATT CTTTACGCGC GCTATCTGTC      300
ATGGTCATCA CTCCTTTTAA CAATTAGGCA GACCAAACGA CATGCATTCTG TCGTATAGCT      360
CTTCATTACT TATGCTTGCC TTATAGTTTT CAATCACATT GCTAACTTCT TTATGACTCA      420
TTGCTTTAAC TTGTTGCTCT GTATATTTTT CGCAGTCTTC TAATTCCAGT TGCTCCTGTA      480
ATGACATCAC ATATTCAACT TGTCTTTGGG TTGCCATCGT TAACCCTCCC ACAAGTCAAA      540
AGCTCTTTGG ACGTAAAACT TCGCCTTTGC TAAATCCTCA TGACCATCTT TTAACGGTGC      600
TCTAGACATG TATTTGATTG CATTACCTAT TGCGAATGCT AGTTGAGGTG GATACTGTGC      660
CGTAACCTGT TCGATAAAAT CTATAATTTT AATGTCGCCG TATGTGTAGT GCGCTGGTTG      720
CTTAACATTG TCTTGCGCTT CGTTCATATC TACTTTTCTG TTAGTGATTA CGCTCATTAT      780
GCTTCACTCC ATTCTTGAA CATTTGGTTA TAAGTGACAT CGAACCAGTA CGGATCACGT      840
GAATGTTTTT GTGGCGTTCC ATCATAAAGC CATGGTCTTA ATCTTCTCTT TCTTTCCTGT      900
TCATATTCCG CTCTCACATT TCGTTGGTAT CGGTTCAAAA TCGCTTTTTT TCTGATTTTT      960
TCTCTCCCTT TTTCTTCATC TTtATtTGA CTCTnCATAT ATTCAACTTC TTCTGTAGAT     1020
nTTGAGTCCT TTCTTCACA CAATAATTCA nCGCCGCGC                               1059

```

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30246 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

	GAAGTAAAAG AAGAATTAAA TTAAACATTA ACAATGGATG AAATTGAATA TGTCGGGACA	60
	ATTGTAGGTC CTGCATATCC ACAACAGGAT ATGTAACTG AGTTAAATGG ATTTGCGCA	120
5	TTAACCAAAA TCGATTGGGA AAACGTAACAT ATCAATAATG AAATTACGGA TATACGCTGG	180
	ATTGATAAAG ATAATGATGC GTTGATTGCG CCTGCTGTCA AAGTTTGGAT TGAAACTTAT	240
10	GGTGGTAAAC ATGACAAATA ATGACACCAT CATGTTACGA CATTATGTCC CACAAGATTA	300
	TTGATGTTA GAAGCTTTTC AATTAAGTGA AAGTGATTG AAGTTTGTTA AAACGCCAGA	360
	GGAAAATATT ACAGCTGCAA TGTCTGATAA TGAAAGGTAT CCCATCGTTG TAATGGATGG	420
15	CAGGCAATGT GTGGCCTTTT TTACATTACA TCGTGGAAAA GGGGTCGCAC CATTTAGCGA	480
	TAACCAAGAT GCAGTATTTT TCAGGTCATT TAGTGTGAT CAACGTTATC GTAATAGAGG	540
	AATAGGTAAA GTGGTAATGG AAAAAATTGGC GTCATTTATC ACTTCAACAT TTCAGGATAT	600
20	TAATGAGATT GTGTAAACGG TTAATACTGA CAATCCACAT GCCATGGCAC TTTATCGCCA	660
	ACAAGGATAT CAATATATGG GAGATAGTAT GTTCGTCGGA AGACCTGTTC ATATTATGGC	720
	GTAACTATA AAATAAATTA AATTTAAAAG CATCTTTACT CATCGTCGAC CACAACAATT	780
25	AATGATGAAT AAAGGTGCTT TTTGTTATAG ATCATCGGAC AATTACTAT AGTAAAAAGC	840
	GACCTAGTGA ACAATTGACA TATATCCACA GGTCGCTTAA CTTAAGTTAT ATTGCTAGTT	900
	GCGATTAATT GATAGACTCA TCATTTTTCG GCTGTGAGA TGGTCTTTT ATTAAAAATG	960
30	CCGTAATCCA AGCCGTAATC GGAATACTGA TTGCAACGGC AATACCGCCT AAAATAATAG	1020
	AAATAAATTC TTGGGCAAT ATTTTCGAGT TTATAATATG ACCAAATGAA TATTTAAGTT	1080
35	TGAAAAACCA AATAAATAAA GCAAGTTGGC CACCAAAAAA GGCAAGGTAA ATCGTGTTG	1140
	CAGATGTCGC TAAAATTTCT CTACCAACAC GCATGCCAGA TTGGAATAAT TCGTATTGCG	1200
	TAACGTTgGA TTCATTGAT GCAATTCATA AATGGGTGAA CTAATGGTAA TTGTAAATC	1260
40	TATCAGCT GCAATAACAG CAAGAATAAT AGTGAACACC ATAAATTGAA CCATATCAAT	1320
	GCCAAATATC ATTGAATACA CATATGTTTC ATCTTGTGT TCGGTTGaAA AGCCTGTAG	1380
	ATGACCGAAG TAGACCGATA AATAAATGAG TGTAATCAAC AATATTGTTG TAACGATAgT	1440
45	GCTGgATAAA TGCaGCTTGT GTTTTAACAT TGTAATATT GAGTACGAAT AAATTACAAG	1500
	CGCCAATAAT AATGCAGAAA AAGAATGTGA CGACATAAAT CGGTACGCCA AAAATAATCA	1560
	ATACAATACT AATAATTAAA ATAGCGAAAT TTAAAAATAG GGTAAATAA GAGATGAATC	1620
50	CCTTTTACC TCCGAAAATT ATCATCAGAA AGAGGAGCAA TAACGCCAAT ATAAATACAG	1680
	CATTCATTGT TTCGCCCTCC TTAATGTTTC AAATATTTCC ATAAACAATA TTGTGATAGG	1740
55		

	CATCGAAATA GTATAAGTCA CTGTATTGGC ATTTTTTAAA AAGATTAAAA ACATAGGTAG	1860
	TGCACCGGAT AAATATGAGA ATAATAAGAT GTTAGTCATT GTTCCATAA TATCTTGGCC	1920
5	GATGTTTCGC CCAGCAAGCG CCCATCTCCT CATTGAAATG TGTGGCGTAC GCTGTAAAT	1980
	TTCATGCATA CCACTAGCAA TTGTAATTGC AACATCCATA ATAGCGCCAA GTGAACCTAT	2040
	TAACACTGAG GCTAGGAAGA TATCTTTCGG TGGTAATGAT AAAAAGTTCA TCGTTTCATA	2100
10	TTTAATGCCT TTACCATCTG TCATATATAT GATTAATTCT GTTAAACCTA TACTCAAAAA	2160
	AGTTCCGATA ATTGTACTGG CTATGGTAAT GAGTGACGC ATATGCCAGC CTGTAACGAG	2220
	CAATAAAGTG AGTATTGTTG AACAGATCAT GGCAATGGTC ATGAGTAAGA ATAAATTAAT	2280
15	ATTGCTATGT TGAATATGAA TGTAAATTGC GATTAATATG GCAATAGAAT TCAAGATTAA	2340
	CGATAAAATC GATTGCAGTC CGACTTTGCG ACCAACCAAT AATACAGTTA ATAAGAACAA	2400
	ACCAGTGATG ATAACCGTTA AGGTATCACG CTTCTTTTCT ATAATATAAG CATCACTCGG	2460
20	CTTGTTAGAA ATATGTAATA ATACTTTTTTCT GTGTGTGCGA AATGCCTCAG AATCTGCTTG	2520
	CGATTTGACG TACTGATGAT TAATCGTCGT CGTTTCTCCA GCAAATTGAC CATTTAATAT	2580
25	TTTGACTTTT AATTGATTTT TATATTTAAT ATCACGATTA TTTTGTGCAT CTTTGTAGG	2640
	TGTCGAAGAA ACATGTTTGA CATCTATAAT TTGACCAATT GGTGTGTGT AAAAGTTCTC	2700
	ATTATTGAAT GTAAATAAAA TAGCACCAAT GAATGCGATG CAGAACAAAC CTAAAATTAT	2760
30	ATTAAATGGC TTTGTAAATA AATTTCTATA TTTCAAAAAC AAAACCCCAA TTCTATGAAT	2820
	GAATTAATAT GGTGATTATA CGCCCTTAAT TTTTTATTTT CAAAGATATT ACTGCTAAGT	2880
	GTAAACGAA AATCATCATT GATAGCATCG AATTACTTAA TGGAATGTAG ACGTTTGTAGT	2940
35	CATTAATTGC TGAATAAGTG TTAATAATAT GCCAATATCA CTCTTTGTAT AAGGCTCCTT	3000
	TGTAATAGCA CATATCGTTC TTTTAAATTC AGTATGATCT AATTTTATAT CTATCCATGA	3060
	TTTAGATTCT GGTAAATGTA TATTTTGTGA TGAAATGATG TAACCTTCTT TTTGACGAAG	3120
40	GAGATACTGC GCAAGTGGTT GGCTACTGAT TGTGTATACA TCTGATTAG TAATCTTGCG	3180
	CAATTGTTTT TTTACAGTTT CGGCAAATGG TGCCAAGCAA TAAATATGAC TATGCTCAAA	3240
	CTGAATTAAT GGTGGGTGTG TCGCCATCGT AATTGGATCG TCTGAAGGCG CATATAAATG	3300
45	ATAGTGCTCT TCGAATAAAG GTAGCATATG TAATTGTTTG TGTTTACGTA TTTCTGGTGT	3360
	AAGTTCCGTG AAACCAATGT CTATATTTCC ATTTAATACG CTATTTATAA TTGTGTCATG	3420
50	TTCTAATAAG CTCGGTATGA CATGTGTATC ATTTTGTAAA TGAAACGTTT GGATAAGTGG	3480
	TAGTAACATG TGGGATACGT CACTCTCATC ATAGCCAATG TAGATACTTT TATTTTGTAGT	3540

55

	TTCAATTAAAT AATAATTTCC CTTCAGATGT GAGCGTAATA TTGCGTCCTT GCTTTTTTAAA	3660
	TAAAGACACA TTAAGTTCTT GTTCTAATAA TGTAATTTGA CGGCTTATCG CTGATTGAGC	3720
5	AATGTTTAGT TCAAGTGCTG TTTCCGAGAT ATGTTCTCTT TTAGCGACCT CGATAAAATA	3780
	TCTTAATTGT TTAATTTCCA TAGCGATATA GGCACCTCCA AAAATGAGTG TTTTGTAACCT	3840
	ATTATAGCAA TATTATTGAT AAATGTTCTA TTTTITAGAT GAATATCTTC TATTTTATAT	3900
10	ATTGAACAGA TAAATTTTTT AGATTATAGT AATTATCATT AATAACTAAT ATCAGAATAT	3960
	TCTAAAAAAG GGGTGTGCAT CATGCACAAT GAGAAATTAA TTAAAGGCTT ATATGACTAT	4020
	CGTGAGGAAC ATGATGCGTG TGGTATTGGT TTTTATGCGA ATATGGATAA TAAAAGGTCT	4080
15	CACGACATCA TTGATAAATC GCTTGAAATG TTGCGACGCT TAGATCACAG GGGCGGGGTC	4140
	GGCGCAGATG GCATCACTGG TGATGGCGCA GGTATTATGA CTGAAATACC TTTTGCATTT	4200
20	TTCAAACAAC ATGTAACGGA CTTTGATATC CCAGGTGAAG GTGAATATGC CGTGGGGTTA	4260
	TTTTTTTCCA AAGAACGCAT TTTAGTTCTT GAACATGAAG TAGTTTTTAA AAAATATTTT	4320
	GAAGGCGAAG GGTATCAAT TCTTGGTTAT CGTAATGTAC CAGTTAATAA AGATGCCATT	4380
25	GCTAAACATG TAGCAGATAC GATGCCAGTC ATTCAACAAG TGTTTATTGA TATTAGGGAC	4440
	ATTGAAGATG TTGAAAAGCG TTTGTTTTTA GCGAGAAAAC AATTAGAGTT CTATTGCACT	4500
	CAGTGCGATT TAGAATTGTA TTTTACGAGC TTATCACGCA AAACAATTGT ATATAAAGGT	4560
30	TGTTTACGAT CAGACCAAAT TAAAAAATA TATACAGATT TATCGGATGA TTTATATCAA	4620
	TCAAAGCTAG GGTTAGTGCA TTCGAGATTT AGTACGAATA CATTCCCGAG TTGGAAAAGG	4680
	GCACATCCTA ACCGTATGTT AATGCATAAT GGTGAGATTA ACACGATTAA AGGTAATGTA	4740
35	AACTGGATGC GAGCACGCCA ACATAAATTA ATCGAAACAT TATTTGGCGA GGATCAACAT	4800
	AAAGTGTTTC AAATTGTGCA TGAGGATGGT AGTGACTCTG CCATTGTAGA TAATGCGCTA	4860
	GAGTTCTTAT CGTTAGCCAT GGAGCCAGAA AAGGCAGCGA TGTTACTCAT ACCTGAACCT	4920
40	TGTTTATATA ATGAAGCGAA TGATGCAAAT GTACGTGCGT TTTATGAATT TTATAGTTAT	4980
	TTAATGGAAC CGTGGGATGG TCCTACAATG ATTTTCGTTCT GTAACGGTGA CAAACTTGGC	5040
	GCGCTTACAG ATAGAAATGG ATTACGTCCA GGTGCTTATA CGATTACTAA AGATAACTTT	5100
45	ATTGTCTTTT CATCTGAAGT GGGTGTGTG GACGTACCTG AAAGTAATGT TGCTTTTAAA	5160
	GGTCAATTGA ATCCTGGAAA GTTATTGCTT GTTGATTTTA AACAGAATAA AGTCATTGAA	5220
50	AATAATGATT TAAAAGGTGC GATTGCTGGA GAATTACCAT ATAAAGCGTG GATTGATAAC	5280
	CATAAAGTTG ACTTTGATTT TGAAAATATA CAATATCAAG ATTCGCAATG GAAAGATGAG	5340

55

	CAGGAACTTG TAGAAGGTAA GAAGGATCCT ATCGGTGCAA TGGGATATGA TCGGCCAATT	5460
5	GCAGTGTGTA ACGAGCGACC AGAATCACTA TTTAATTACT TTAAACAGCT GTTTGCACAA	5520
	GTTACGAATC CACCAATTGA TCGGTATCGT GAAAAAATCG TAACGAGTGA ACTTTCTTAT	5580
	TTAGGTGGCG AAGGTAACCT ACTAGCACCT GACGAAACGG TTTTAGATCG TATTCAATTG	5640
10	AAAAGGCCGG TATTGAATGA ATCACACTTA GCAGCGATTG ATCAGGAACA TTTTAAATTA	5700
	ACTTATTTAT CAACGGTATA TGAAGGGGAT TTGGAAGATG CGTTAGAAGC ATTAGGCCGA	5760
	GAAGCAGTGA ATGCTGTAAA GCAAGGCGCT CAAATTCTAG TGTAGATGA TAGTGGATTA	5820
15	GTTGATAGCA ATGGCTTTGC AATGCCGATG TTAATCGCAA TAAGTCATGT GCATCAATTA	5880
	CTTATTAAAG CAGATTTACG TATGTCTACA AGTTTAGTCG CTAAATCTGG TGAGACACGA	5940
	GAAGTGCATC ATGTTGCTTG TTTACTCGCA TATGGCGCGA ATGCAATTGT GCCATACCTA	6000
20	GCGCAACGTA CAGTTGAACA ACTGACATTG ACAGAAGGGT TACAAGGCAC CGTTGTCGAT	6060
	AATGTTAAGA CATATACGGA TGTATTGTCA GAAGGTGTCA TTAAAGTAAT GGCTAAGATG	6120
	GGAATTTTCA CAGTGCAAAG TTATCAAGGG GCACAAATAT TTGAAGCGAT TGGCTTGTCT	6180
25	CATGATGTGA TTGATCGTTA TTTTACTGGG ACACAGTCTA AGTTATCTGG TATTTGATTT	6240
	GATCAAATTG ATGCTGAAAA TAAAGCACGT CAACAAAGTG ATGATAATTA TCTTGATCA	6300
	GGTAGTACAT TCCAATGGAG ACAACAAGGT CAACATCATG CTTTTAATCC GGAATCTATT	6360
30	TTCTTATTGC AGCACGCATG TAAAGAAAAT GACTATGCGC AATTTAAAGC ATACTCTGAA	6420
	GCGGTGAACA AAAATAGAAC AGATCACATT AGACATTTAC TTGAATTTAA AGCATGTACA	6480
	CCGATTGACA TCGACCAAGT TGAACCGGTA AGTGACATTG TCAAACGCTT TAATACAGGG	6540
35	GCGATGAGTT ATGGATCGAT TTCAGCGGAA GCACATGAAA CGTTAGCACA AGCCATGAAC	6600
	CAATTAGGTG GAAAGAGTAA TAGTGGTGAA GGTGGCGAAG ATGCAAAACG TTATGAAGTA	6660
40	CAAGTTGATG GAAGCAACAA AGTAAGTGCG ATTAAACAAG TTGCTTCTGG GCGTTTTGGT	6720
	GTAAGTAGTG ATTATTTTACA ACATGCCAAA GAAATTCAAA TTAAAGTTGC GCAAGGTGCA	6780
	AAGCCTGGTG AAGGTGGTCA ATTACCTGGT ACTAAGGTAT ATCCGTGGAT TCGAAGACA	6840
45	AGAGGGTCAA CGCCAGGTAT CGGTCTGATT TCACCACCGC CACATCATGA TATTTATTCA	6900
	ATAGAAGATT TAGCGCAACT GATACATGAT TTGAAAAATG CGAATAAAGA TGCAGATATC	6960
	GCGGTAAAT TAGTTTCGAA AACAGGTGTT GGTACCATTG CATCTGGGGT GGCAAAAGCA	7020
50	TTTGACAGATA AAATTGTTCAT CAGTGGTTAC GATGGTGGTA CAGGGGCTTC ACCCAAAACG	7080
	AGTATTCAGC ATGCCGGTGT TCCTTGGGAG ATTGGTTTAG CAGAAACACA TCAAACATTA	7140

55

	AAAGATGTAG CGTACGCATG TGCCTTGA GCGGAAGAAT TTGGATTGTC AACTGCACCA	7260
	TTAGTGGTGT TGGGCTGTAT TATGATGCGT GTATGCCATA AAGATACATG TCCAGTAGGA	7320
5	GTTGCAACTC AAAACAAAGA TTTACGTGCT TTATATAGAG GTAAAGCACA TCATGTTGTT	7380
	AATTTTATGC ATTTTATTGC ACAAGAATTA AGAGAAATTT TAGCATCTTT AGGTTTGAAA	7440
	CGTGTAGAAG ACTTAGTTGG AAGAACTGAT TTATTACAAC GATCATCAAC ATTAAGGCG	7500
10	AATAGCAAAG CGGCTAGTAT TGATGTTGAA AACTGTTAT GTCCTTTCGA TGGGCCAAAC	7560
	ACAAAAGAAA TTCAACAAAA TCATAATCTT GAGCATGGAT TTGATTTAAC AAATTTATAT	7620
	GAAGTAACGA AGCCATATAT TGCTGAAGGG CGTCGCTATA CAGGTAGCTT TACAGTAAAT	7680
15	AATGAACAAC GTGATGTAGG GGTATTACA GGTAGTGAGA TTTCGAAACA ATATGGAGAA	7740
	GCAGGACTTC CTGAAAATAC AATTAATGTT TATACGAATG GTCATGCTGG TCAAAGTCTT	7800
	GCAGCATATG CACCGAAAGG CTTAATGATT CATCATACTG GAGATGCGAA TGAATATGTT	7860
20	GGTAAAGGAT TATCTGGTGG TACGGTCATT GTCAAAGCAC CTTTGAAGA ACGACAAAT	7920
	GAAATTATTG CTGGTAACGT CTCATTCTAT GGTGCGACAA GTGGTAAGGC ATTTATTAAC	7980
25	GGTAGTGACG GAGAAAGATT CTGTATTAGA AATAGTGGTG TAGATGTTGT CGTTGAAGGT	8040
	ATCGGCGACC ATGGATTAGA GTATATGACT GGTGGACATG TCATTAATTT AGGTGATGTA	8100
	GGTAAGAACT TCGGTCAAGG TATGAGTGGT GGTATTGCTT ACGTTATCCC GTCTGATGTA	8160
30	GAAGCTTTTG TTGAAATAA TCAACTAGAT ACGCTTTCGT TTACAAAGAT TAAACACCAA	8220
	GAAGAAAAAG CATTCAATTA GCAATGCTG GAAGAACATG TGTCACACAC GAATAGTACG	8280
	AGAGCGATTG ATGTGTTAAA ACATTTTGAT CGCATTGAAG ATGTCGTCGT TAAAGTTATT	8340
35	CCTAAAGATT ATCAATTAAT GATGCAAAAA ATTCATTTGC ACAAATCATT ACATGACAAT	8400
	GAAGATGAAG CGATGTTAGC TGCATTTTAC GATGACAGTA AAACAATCGA TGCTAAACAT	8460
	AAACCAGCCG TTGTGTATTA AGGAAAGGGG GAGATACGAT GGGTGAATTT AAAGGATTTA	8520
40	TGAAGTATGA CAAACAGTAC TTAGGTGAAT TATCACTGGT AGACCGTTTG AAGCATCATA	8580
	AAGCATATCA ACAACGATTT ACTAAAGAAG ATGCCTCTAT CCAAGGTGCA CGATGTATGG	8640
	ATTGTGGAAC GCCGTTTGT CAAACCGGAC AACAGTATGG TAGGGAAACA ATAGGTTGTC	8700
45	CAATTGGAAA CTACATTCCT GAATGGAACG ACTTAGTGTA TCATCAAGAT TTTAAACTG	8760
	CTTATGAACG CTTAAGCGAA ACAAATAACT TTCCTGACTT TACAGGGCGT GTATGTCCTG	8820
50	CACCATGCGA AAGTGCTTGT GTGATGAAGA TTAATAGAGA ATCGATTGCG ATTAAAGGTA	8880
	TTGAACGCAC AATTATTGAT GAAGCTTTTG AAAATGGTTG GGTAGCGCCG AAAGTCCGA	8940

55

	CTGAAGAACT TAATCTACTA GGATATCAAG TAACTATTTA TGAACGTGCT ACAGAATCAG	9060
	GCGGTTTATT AATGTATGGT ATTCCGAATA TGAAACTTGA TAAAGATGTG GTTCGACGTC	9120
5	GTATTAAGTT AATGGAAGAA GCGGGCATT CTTTCATTAA TGGTGTGAA GTCGGTGTG	9180
	ATATTGATAA AGCAACGTTA GAATCTGAGT ATGATGCCAT TATATTATGT ACTGGTGCAC	9240
	AAAAAGGTAG AGATTTACCT TTAGAAGGAC GCATGGGTGA TGGTATACAT TTCGCTATGG	9300
10	ATTATTTAAC TGAACAAACG CAGTTGTTAA ATGGAGAAAT TGATGATATA ACAATAACTG	9360
	CAAAAGATAA GAATGTCATT ATCATTGGTG CTGGTGATAC AGGGGCAGAC TGTGTAGCGA	9420
	CAGCATTAAAG AGAAAATTGT AAATCGATTG TTCAATTTAA TAAATATACG AAATTGCCAG	9480
15	AAGCAATTAC ATTTACAGAA AATGCATCAT GGCCTTTAGC AATGCCGGTG TTTAAAATGG	9540
	ACTATGCGCA CCAAGAGTAC GAAGCTAAGT TTGGTAAGGA ACCACGTGCA TATGGTGTTC	9600
20	AAACAATGCG TTACGATGTT GACGATAAAG GACACATACG TGGTTTGTAT ACTCAAATTT	9660
	TAGAGCAAGG CGAAAATGGT ATGGTCATGA AAGAAGGACC TGAAAGATTT TGGCCTGCTG	9720
	ACCTTGATT ATTATCAATC GGCTTCGAAG GTACAGAACC AACAGTACCG AATGCTTTTA	9780
25	ACATTAAAC GGATAGAAAT CGAATCGTGG CGGATGATAC AACTATCAA ACTAATAATG	9840
	AAAAGGTATT TGCTGCTGGA GATGCTAGAC GTGGTCAAAG TTTAGTTGTA TGGGCAATTA	9900
	AAGAAGGTAG AGGCGTAGCG AAAGCAGTAG ATCAGTATTT AGCTAGTAAA GTTTGTGTAT	9960
30	AATCTTTGTA TGGAATGGT GGTTACGTTG ACGTTGTGAC ATGCTGAATC GAGTTTGAAA	10020
	AAATCTAGTA TCTATCAACG TCACATGCCA TCTTTGTAAC CTAAAAACAA AGGTTTGTA	10080
	GACAACAAAT AGATTAATTA TAAGTAGTGA TTTTITACAT TCGTTTATAG GTCAACTGTA	10140
35	GTGGAAGACA ATGATTTGTG GTAATCATGT AATGCTTAAA AACAAATATTG ACTTTTACAG	10200
	AACGTTTATA TATGATAAAT ATTGTGTTTA GGAGGAATAC CCAAGTCCGG CTGAAGGGAT	10260
	CGGTCTTGAA AACCGACAGG GGCTTAACGG CTCGCGGGG TTCGAATCCC TCTTCCTCCG	10320
40	CCATCAATAT TTATATTAAA TTCTATATAT AATGAAGGTA AGTGCTCAA TTTTGAGTAT	10380
	TTACCTTTTT TATTTGTCTT TGAATGGCTC GTAATTTTGA ATAATAGAAA TGATAAGGCA	10440
	TTGAGATTGG AAGGGCATT GGCTTGTCGA ATATACATAG CTAAATGTCT TTTTGTGTTT	10500
45	GTGAAATATG ATGGATGGCT TGTGTGGACA AGTTTGCTAT TTATAGATAT GCATTTTCA	10560
	ATTTAGGAGT TGGCCATGCA TCTACACTTT ATAATGGTGA GAGCGTGGTG AGGTATTGTT	10620
50	AATAACGCAA TTGTAGCGAG GAGTTATTGC TACATATGTC GTTATGGCTC ATTGATTTTC	10680
	TGAAATGGCT ACCCCAGATA ATTGTGACAA AATAAAAATA TTTTGTTGAA AGCCTTTACA	10740

55

	TAAAAAGAGA AGATGTAAAA GCCATCGTAA CCGCTATTGG GGGAAAAGAA AATCTTGAAG	10860
	CTGCAACGCA TTGTGTAACA CGATTACGTT TAGTGCTGAA GGATGAAAGT AAAGTTGATA	10920
5	AAGACGCATT AAGTAATAAC GCGTTGGTCA AGGGGCAGTT TAAAGCAGAC CATCAATATC	10980
	AAATTGTCAT TGGTCCAGGA ACAGTCGATG AAGTGATATA GCAGTTTATT GATGAAACAG	11040
	GTGCTCAAGA AGCTTCGAAA GATGAAGCGA AACAAGCAGC TGCACAAAAA GGAATCCAG	11100
10	TACAACGTTT GATCAAATTG TtGGGGGATA TTTTATAACC AATATTACCT GCGATTGTGA	11160
	CAGCTGGTTT GTTAATGGGA ATCAATAATT TACTTACAAT GAAAGGTTTA TTTGGTCCAA	11220
	AAGCACTTAT TGAGATGTAT CCACAAATTG CTGATATTTT AACATCATT AATGTGATTG	11280
15	CGAGTACGGC ATTTATTTTC TTACCAGCAT TAATTGGTTG GAGTAGTATG CGTGTATTG	11340
	GTGGTAGTCC GATTCTAGGC ATAGTCTTAG GTTTGATTTT AATGCATCCG CAATTAGTAT	11400
	CTCAGTATGA TTTGGCAAAA GGAATATTC CGACGTGGAA CTTATTTGGC TTAGAGATTA	11460
20	AGCAGTTGAA TTACCAAGGT CAAGTGTTGC CAGTtTTAAT TGCAGCTTAC GTTCTAGCTA	11520
	AAATTGAAAA AGGATTAAAT AAAGTCGTTT ACGATTCGAT AAAAATGTTG GTCGTTGGAC	11580
25	CCGTAGCGCT TTTAGTACT GGATTTTTAG CATTTATTAT CATTGGACCA GTTGCGTTAT	11640
	TGaTTGGTAC AGGTATTACA TCTGGTGTTA CATTTATATT CCAACATGCA GGATGGCTTG	11700
	GCGGAGCAAT ATATGGATTG TTATATGCAC CACTTGTAAT TACAGGACTA CACCATATGT	11760
30	TTTTAGCAGT AGATTTCOA TTGATGGGTA GCAGCTTAGG CCGTACGTAT TTATGGCCAA	11820
	TTGTTGCGAT TTCCAATATT TGTGAGGGCT CTGCAGCATT TGGAGCATGG TTTGTCTATA	11880
	AACGTCGTAA AATGGTTAAA GAAGAAGGCT TGGCATTAA ATCTTGTATT TCTGGTATGT	11940
35	TAGGTGTTAC TGAACCAGCC ATGTTGGTG TGAACCTACC TCTGAAATAT CCATTTATCG	12000
	CTGCGATATC AACGTCTTGT GTATTGGGGG CAATCGTTGG TATGAATAAC GTACTTGGA	12060
	AAGTTGGTGT TGGTGGCGTG CCAGCATTCA TTTCAATTCA AAAAGAATTT TGGCCAGTAT	12120
40	ATCTTATTGT GACAGCTATT GCTATTGTTG TACCATGTAT ACTAACAATT GTGATGTCTC	12180
	ATTTTAGTAA ACAAAAAGCG AAAGAAATTG TTGAAGATTA ATAAAAATAA AAAGGGGCGT	12240
	TCGTTATTTG GACGTCCTTT ATTACGTTAT AAGGTGGTAA TTGTGTGTCG AAAGAAATAG	12300
45	ATTGGAGAAA ATCCGTTGTA TATCAAATTT ATCCTAAGTC GTTTAATGAT ACGACGGGGA	12360
	ATGGTATAGG AGATATCAAT GGAATTATAG AAAAATTGGA TTATATCAAG TTATTGGGTG	12420
50	TTGATTATAT TTGGTTAACA CCAGTGATG AATCACCGAT GAATGATAAT GGCTATGATA	12480
	TCAGCAATTA TTTAGAAATC aATGAAGACT TTGGAACGAT GGATGATTTT GaAAAGTTAA	12540

55

	CGACGGAGCA	TGaATGGTTT	AAAGAAGCCC	GTAAATCTAA	AGATAACCCy	TATAGAGATT	12660
	ATTACTTTTT	CAGATCATCT	GAAGACGGGC	CGCCAACAAA	TTGGCATTCT	AAATTCCGGT	12720
5	GTAATGCATG	GAAGTATGAT	TCTGAGACAG	ATGAATATTA	TTTACATTTA	TTTGATGTCA	12780
	GTCAAGCTGA	TTTAAATTGG	GATAATCCGG	AAGTACGTCA	ATCGTTATAT	CGCATAGTCA	12840
	ATCATTGGAT	AGACTTCGGC	GTTGATGGTT	TTCGATTGGA	TGTCATTAAC	TTAATTTCTA	12900
10	AAGGTGAATT	TAAGGACTCT	GACAAAATAG	GTAAAGAATT	TTATACGGAT	GGTCCTAGAG	12960
	TGCATGAGTT	TCTGCATGAA	TTAAATCGTC	AAACGTTTGG	TAACACTGAC	ATGATGACTA	13020
	TAGGAGAAAT	GTCTTCGACG	ACGATTGAAA	ATTGTATTAA	GTATACACAA	CCAGAACGCC	13080
15	AAGAATTGAA	TAGTGTTTTT	AATTTTCATC	ATCTAAAGGT	TGATTATGTT	GATGGTGAAA	13140
	AGTGGACAAA	TGCGAgcTTG	nATTTTCATA	AGTTAAAGGA	AATTCTGATG	CAATGGCAAC	13200
	GAGGTATTTA	TGACGGTGGC	GGATGGAACG	CGATTTTCTG	GTGTAATCAT	GATCAGCCAC	13260
20	GGGTAGTGTC	TAGATTTGGT	GATGATACGT	CGGAAGAGAT	GAGGATACAA	AGTGCTAAAA	13320
	TGTTAGCTAT	CGCACTGCAT	ATGTTGCAAG	GGACGCCATA	TATTTACCAA	GGTGAAGAAA	13380
	TTGGTATGAC	GGACCCACAT	TTTACATCAA	TAGCACAATA	TCGTGATGTT	GAATCGATTA	13440
25	ATGCCTACCA	TCAGTTGTTA	AGTGAAGGGC	ATGCTGAAGC	GGATGTGTTA	GCGATTTTAG	13500
	GACAGAAGTC	ACGAGACAAT	TCGAGAACGC	CTATGCAATG	GAGTGATGAT	GTTAATGCTG	13560
30	GATTTACAGC	TGGTAAnCCT	TGGATTGATA	TTTCGGAAAA	TTATCATCAG	GTCAACGTTA	13620
	GACAAGCACT	TCAGAATAAA	GAGTCTATTT	TCTATACGTA	TCAAAAATTA	ATACAATTAA	13680
	GACATACGCA	TGATATTATT	ACGTATGGAG	ACATTGTGCC	ACGTTTTATG	GATCATGATC	13740
35	ATTTATTTGT	TTATGAACGT	CATTATAAGA	ATCAACAATG	GCTAGTAATT	GCGAATTTCT	13800
	CAGCAtCGGC	TGTTGATTTG	CCAGAAGGAT	TGGCTAGAGA	AGGTTGTGTT	GTGATTCAAA	13860
	CAGGCACAGT	GGAAAATAAT	ACGATAAGCG	GGTTTGGTGC	AATTGTAATC	GAAACAAACG	13920
40	CGTAAAATAA	ATTGAGTGGA	TGCGTTTATA	TGGCGAAACA	AAAAAAGTTT	ATGAAGATTT	13980
	ATGAGGCGTT	GAAAGAAGAT	ATATTAAACG	GGCAGATTCA	ATATGGTGAA	CAAATTCCGT	14040
	CTGAACATGA	TTTGGTGCAA	TTGTACCAGT	CATCTCGAGA	GACCGTGCGT	AAGGCATTAG	14100
45	ATTTGTTGGC	ATTAGACGGC	ATGATTCAAA	AGATTCATGG	TAAAGGGTCA	CTTGTCATTT	14160
	ATCAGGAGGT	TACAGAGTTT	CCATTTTCTG	AACTTGTTAG	TTTTAAAGAA	ATGCAAGAAG	14220
	AAATGGGCGT	CGCATATTTA	ACTGAAGTTG	TTGTGAATGA	GGTTGTTGAA	GCGCATGAAG	14280
50	TTCCAGAAGT	TCAACATGCT	TTAAACATCA	ATTCTAGTGA	ATCACTCATT	CATATTGTTA	14340

55

	TTGTTTCAGA TATAGGTAAT GATGTTGCGA GTGATTCTAT TTATGATTAT TTGGAAAAGG	14460
	TATTAAATCT TAATATTAGT TATTCAAGTA AGTCTATTAC TTTTGAACCG TTTGATGAAC	14520
5	AAGCATATCA ATTGTTTGGT GATGTATCGG TGGCTTATTC AGCAACAGTT CGAAGTATTG	14580
	TGTATTTAGA AAATACAATG CCGTTTCAAT ATAATATTTT AAAACATCTT GCAAATGAAT	14640
	TTAAATTTAA TGACTTCTCA AGACGTCGTA TAAAGTAAAC AATGATATAA ATGATTTATA	14700
10	CTTGCAATTA ACTATTAAAA TATAGTAATA TATATCTTGC CGTGCTAGGT GGGGAGGTAG	14760
	CGGTTCCCTG TACTCGAAAT CCGCTTTATG CGAGGCTTAA TTCCTTTGTT GAGGCCGTAT	14820
	TTTTGCGAAG TCTGCCCAA GCACGTAGTG TTTGAAGATT TCGGTCCTAT GCAATATGAA	14880
15	CCCATGAACC ATGTCAGGTC CTGACGGAAG CAGCATTAAAG TGGATCATCA TATGTGCCGT	14940
	AGGgTAGCCG AGATTTAGCT AACGACTTTG GTTACGTTTCG TGAATTACGT TCGATGCTTA	15000
	GGTGCACGGT TTTTTATTTT TTAAATATTA AACCGATTAT TAAGAGTTGA AAATATATAA	15060
20	TTATAGAAGC TACTTTCTTG AAGACAATTC AGCGTATTAT ACGTGGAACA TGTTTGTGGG	15120
	AAGTAGCTTT TTTATATGTG AAGTTTGATT CAAGTGAAC TCGATGTGCAG TTTGAATGAT	15180
	TTTTGTGTCA ATGAAAAGTA AGAAGTTATA ATTTGATGAT AAAGAAATGA TGGTGAAATG	15240
25	AGGGGGAGTA TCTTACAATA GAATTATTAA TGAGATACGT TATGATTATT GACAATCAAA	15300
	TGCCTACGGA GGACATATGC AAATATATTT AAGTACTTTA ACAGAGTTAG ATTATGATAA	15360
30	ATCTTTAAAT AGTATTGAAG AAAGTTTGA TGATAATCCT GAAACGAGTT GGCAAGCACG	15420
	TGCGAAAGTA AAACATTTAA GAAAATCTCC TTGCTATAAT TTTGAATTAG AAGTAATAGC	15480
	GAAAAATGAA AATAACGATG TCGTTGGACA CGTTTTATTA ATTGAAGTAG AAATTAATAG	15540
35	TGATGATAAG ACGTATTATG GTTTGGCGAT TGCCTCTTTA TCAGTTCATC CTGAATTACG	15600
	TGGA C AAAAA TTAGGTCGTG GCTTGGTTCA AGCAGTAGAA GAGCGTGCCA AAGCACAAGA	15660
	GTATAGTACG GTTGTTGTAG ACCATTGTTT TGA C TACTTT GAAAAGTTGG GTTATCAAAA	15720
40	TGCTGCTGAG CATGACATTA AATTAGAATC TGGTGATGCA CCGTTACTTG TAAAATATTT	15780
	ATGGGATAAT TTGACGGATG CACCACACGG AATCGTAAAA TTTCCAGAAC ATTTTTATTA	15840
	ATTGTTCAAT TAAGAAGTAA AGGTATTATC ATGCTATAAT GAGAGGTAAT TGTTTATGGA	15900
45	GGTGCTAACT TGAATTATCA AGCCTTATAT CGTATGTACA GACCCCAAAG TTTCGAGGAT	15960
	GTCGTCGGAC AAGAACATGT CACGAAGACA TTGCGCAATG CGATTTTCGAA AGAAAAACAG	16020
	TCGCATGCTT ATATTTTATG TGGTCCGAGA GGTACGGGGA AAACGAGTAT TGCCAAAGTG	16080
50	TTTG c TAAAG CAATCAACTG TCTAAATAGC ACTGATGGAG AACCTTGTA TGAATGTCAT	16140

AATAATGGCG TTGATGAAAT AAGAAATATT AGAGACAAAG TTAAATATGC ACCAAGTGAA 16260
 TCGAAATATA AAGTTTATAT TATAGATGAG GTGCACATGC TAACAACAGG TGCTTTTAAT 16320
 5 GCCCTTTTAA AGACGTTAGA AGAACCTCCA GCACACGCTA TTTTATATT GGCAACGACA 16380
 GAACCACATA AAATCCCTCC AACAAATCATT TCTAGGGCAC AACGTTTTGA TTTTAAAGCA 16440
 ATTAGCCTAG ATCAAATTGT TGAACGTTTA AAATTGTAG CAGATGCACA ACAAATTGAA 16500
 10 TGTGAAGATG AAGCCTTGGC ATTTAtcgCT AAAGCGTCTG AAGGGGGTAT GCGTGATGCA 16560
 TTAAGTATTA TGGATCAGGC TATTGCATTT GGTGATGGTA CGTTAACATT GCAAGATGCG 16620
 TTGAATGTCA CAGGTAGCGT ACATGATGAA GCGTTGGATC ACTTGTTTGA TGATATTGTA 16680
 15 CAAGGTGACG TACAAGCATC TTTTAAAAA TACCATCAGT TTATAACAGA AGGTAAAGAA 16740
 GTGAATCGCC TAATAAATGa TATGATTTAT TTTGTCaGAG ATACGATTAT GAATAAAACA 16800
 TCTGAGAAAG ATACTGAGTA TCGAGCACTG ATGAACTTAG AATTAGATAT GTTATATCAA 16860
 20 ATGATTGATC TTATTAATGA TACATTAGTG TCGATTCGTT TTAGTGTGAA TCAAAACGTT 16920
 CATTTTGAAG TGTTGTTAGT AAAATTAGCT GAGCAGATTA AGGGTCAACC ACAAGTGATT 16980
 GCGAATGTAG CTGAACCAGC ACAAATTGCT TCATCGCCAA ACACAGATGT ATTGTTGCAA 17040
 25 CGTATGGAAC AGTTAGAGCA AGAACTAAAA ACACTAAAAG CACAAGGAGT GAGTGTGCT 17100
 CCTGTTCAAA AATCTTCGAA AAAGCCTGCG AGAGGCATAC AAAAATCTAA AAATGCATTT 17160
 TCAATGCAAC AAATTGCAAA AGTGCTAGAT AAAGCGAATA AGGCAGATAT CAAATTGTTG 17220
 30 AAAGATCATT GGCAAGAAGT GATTGATCAT GCCAAAAATA ATGATAAAAA ATCACTCGTT 17280
 AGTTTATTGC AAAATTCGGA ACCTGTGGCG GCAAGTGAAG ATCACGTACT TGTGAAATTT 17340
 GAGGAAGAGA TCCATTGTGA AATCGTCAAT AAAGACGACG AGAAACGTAG TAGTATAGAA 17400
 35 AGTGTGTAT GTAATATCGT TAATAAAAAAC GTTAAAGTTG TTGGTGTACC ATCAGATCAA 17460
 TGGCAAAGAG TTCGAACGGA ATATTTACAA AATCGTAAAA ACGAAGGCGA TGATATGCCA 17520
 40 AAGCAACAAG CACAACAAAC AGATATTGCT CAAAAAGCAA AAGATCTTTT CGGTGAAGAA 17580
 ACTGTACATG TGATAGATGA AGAGTGATAC ATGACAAGCG ATATAATCGT ATGTATAATG 17640
 AAAGAAACAT CATTTTATTG ATAAATATTT ATTGATTTTC AAGGAGGAAA TGGAAATATGC 17700
 45 GCGGTGGCGG AAACATGCAA CAAATGATGA AACAAATGCA AAAAATGCAA AAGAAAATGG 17760
 CTCAAGAACA AGAAAACTT AAAGAAGAGC GTATTGTAGG AACAGCTGGC GGTGGCATGG 17820
 TTGCAGTTAC TGTAACCTGGT CATAAAGAAG TTGTCGACGT TGAAATCAAA GAAGAAGCTG 17880
 50 TAGACCCAGA CGATATTGAA ATGCTACAAG ACTTAGTGTT AGCAGCTACT AATGAAGCGA 17940

55

	TCCCTGGaAT GTGATCATAG ATGCATTATC CAGAACCTAT ATCAAACTT ATTGATAGCT	18060
	TTATGAAATT GCCAGGCATT GGTCCAAAGA CAGCCCAACG TCTGGCTTTT CATACCTTAG	18120
5	ATATGAAAGA AGACGATGTT GTTCAGTTTG CCAAAGCATT AGTAGATGTT AAGAGAGAAT	18180
	TAACATATTG TAGCGTATGT GGTACATTA CTGAAAATGA TCCATGTIAT ATTTGTGAAG	18240
	ATAAGCAAAG AGATCGTTCA GTTATTTGTG TTGTGGAAGA TGACAAAGAT GTCATAGCTA	18300
10	TGGAAAAAAT GAGAGAATAC AAAGGTTTAT ATCACGTTTT ACATGGGTCT ATTTGCGCTA	18360
	TGGATGGCAT TGGACCAGAA GATATTAATA TTCCTTCATT GATTGAACGC TTGAAAAACG	18420
	ATGAAGTTAG CGAATTAATC TTAGCTATGA ACCCGAACTT AGAGGGGGAA TCTACAGCCA	18480
15	TGTATATTTT TAGATTAGTT AAGCCTATAG GTATCAAAGT GACGAGATTA GCACAAGGGT	18540
	TATCGGTAGG TGGCGATTGA GAGTATGCTG ACGAAGTAAC ATTATCTAAA GCAATCGCAG	18600
	GTAGAACAGA AATGTAATkt CTCTATTAA ACATTTTGA TTTTAATACT ATAGTAAGAA	18660
20	AAGTCACAGT GTAATCATTG TGGCTTTTTT TATGGTGTGG TGTGATGTAC TACTTTATTT	18720
	GCGGTGTGGC GGTGGTATGG TTTACCTAGT TTTACTGAGG GATGGGTAAT CTTTAGGAAG	18780
	CAAGCCGTTG GTTGTGATTT GTTACTTCTA ATAGTAATGA TGTGAATTGG ATTATCGAAT	18840
25	TAGATCTATG GTTATGGTGT GTTGGTGCTA TTAATTTGAT AAATGCGGTT AATGACTATG	18900
	CAAATGAAAT TCTTTTGTA TGAATGAT AGATGCTGGC TTAGTAAGTT GTACTTCTTT	18960
30	GGTCTAAAGC TTATTAAATC AGCCTGTATA GCGGTGTTTT GAGAGATTAT TTAAAACTTG	19020
	TAAATTTATT TTTAATTTCT GGTAAAAAAA TAACGTTCTG TTTTGCCTTT TTTTGTATTG	19080
	ATATGGTTAG AGAAAAATCT GTTCTTGTT CTAAAAACG TACTATTTAT AAGTGGGGAT	19140
35	TTTTTAAGTT CGATTTTATG GATAAGGGCG TTCAGTACAG ATGACAAAGG TGTAATTTTT	19200
	ACTGTGTGTA AGCAGTTTGA AAGCCTGTAT AGTATTTATT TGTGAGGCA AACAAAACAA	19260
	CTCAACTTAA GAAATAACTT GAATTACTAA CGAAAATTAA TTTTAAAAAG TTATTGACTT	19320
40	AAATGTAAAT AAAATGTATA ATTAATTCTT GTCGGTAAGA AAAATGAACA TTGAAAAC TG	19380
	AATGACAATA TGTCAACGTT AATTCCAAAA AACGTAAC TAAGTTACAA ACATTATTTA	19440
	GTATTTATGA GCTAATCAAA CATCATAATT TTTATGGAGA GTTTGATCCT GGCTCAGGAT	19500
45	GAACGCTGGC GGCGTGCCTA ATACATGCAA GTCGAGCGAA CGGACGAGAA GCTTGCTTCT	19560
	CTGATGTTAG CGGCGGACGG GTGAGTAACA CGTGGATAAC CTACCTATAA GACTGGGATA	19620
	ACTTCGGGAA ACCGKAGCTA ATACCGGATA ATATTTTGAA CCGCATGGTT CAAAAGTGAA	19680
50	AGACGGTCTT GCTGTCACTT ATAGATGGAT CCGCGCTGCA TTAGCTAGTT GGTAAGGTAA	19740

55

	GAGACACGGT	CCAGACTCCT	ACGGGAGGCA	GCAGTAGGGA	ATCTTCCGCA	ATGGGCGAAA	19860
	gCtGaCGGAG	CAACGCCGCG	TGAGTGATGA	AGGTCTTCGG	ATCGTAAAAC	TCTGTTATTA	19920
5	GGGAAGAACA	TATGTGTAAG	TAACTGTGCA	CATCTTGACG	GTACCTAATC	AGAAAGCCAC	19980
	GGCTAACTAC	GTGCCAGCAG	CCGCGGTAAT	ACGTAGGTGG	CAAGCGTTAT	CCGGAATTAT	20040
	TGGGCGTAAA	GCGCGCGTAG	GCGGTTTTTT	AAGTCTGATG	TGAAAGCCCA	CGGCTCAACC	20100
10	GTGGAGGGTC	ATTGGAAACT	GGAAAACCTG	AGTGCAGAAG	AGGAAAGTGG	AATTCCATGT	20160
	GTAGCGGTGA	AATGCGCAGA	GATATGGAGG	AACACCAGTG	GCGAAGGCGA	CTTTCTGGTC	20220
	TGTAAGTAC	GCTGATGTGC	GAAAgCGTGG	GGATCAAACA	GGATTAGATA	CCCTGGTAGT	20280
15	CCACGCCGTA	AACGATGAGT	GCTAAGTGTT	AGGGGGTTTC	CGCCCTTAG	TGCTGCAGCT	20340
	AACGCATTAA	GCACTCCGCC	TGGGGACTAC	GACCCGAAGt	TGAAACTCAA	AGGAATTGAC	20400
	GGGGACCCGC	ACAAGCGGTG	GAGCATGTGG	TTTAATTGCA	AGCAACGCGA	AGAACCTTAC	20460
20	CAAATCTTGA	CATCCTTTGA	CAACTCTAGA	GATAGAGCCT	TCCCCTTCGG	GGGACAAAGT	20520
	GACAGGTGGT	GCATGTTGT	CGTCAGCTCG	TGTCGTGAGA	TGTTGGGTTA	AGTCCCGCAA	20580
	CGAGCGCAAC	CCTTAAGCTT	AGTTGCCATC	ATTAAGTTGG	GCACTCTAAG	TTGACTGCCG	20640
25	GTGACAAACC	GGAGGAAGGT	GGGGATGACG	TCAAATCATC	ATGCCCTTA	TGATTTGGGC	20700
	TACACACGTG	CTACAATGGA	CAATACAAAG	GGCAGCGAAA	CCGCGAGGTC	AAGCAAATCC	20760
30	CATAAAGTTG	TTCTCAGTTC	GGATTGTAGT	CTGCAACTCG	ACTACATGAA	GCTGGAATCG	20820
	CTAGTAATCG	TAGATCAGCA	TGCTACGGTG	AATACGTTCC	CGGGTCTTGT	ACACACCGCC	20880
	CGTCACACCA	CGAGAGTTTG	TAACACCCGA	AGCCGGTGGA	GTAACCTTTT	AGGAGCTAGC	20940
35	CGTCGAAGGT	GGGACAAATG	ATTGGGGTGA	AGTCGTAACA	AGGTAGCCGT	ATCGGAAGGT	21000
	GCGGCTGGAT	CACCTCCTTT	CTAAGGATAT	ATTCCGAACA	TCTTCTTCAG	AAGATGCGGA	21060
	ATAACGTGAC	ATATTGTATT	CAGTTTGTAA	TGTTTATTTA	ACATTCAAAT	ATTTTTTGGT	21120
40	TAAAGTGATA	TTGCTTATGA	AAATAAAGCA	GTATGCGAGC	GCTTGACTAA	AAAGAAATTG	21180
	TACATTGAAA	ACTAGATAAG	TAAGTAAAAT	ATAGATTTTA	CCAAGCAAAA	CCGAGTGAAT	21240
	AAAGAGTTTT	AAATAAGCTT	GAATTCATAA	GAAATAATCG	CTAGTGTTTCG	AAAGAACACT	21300
45	CACAAGATTA	ATAACGCGTT	TAAATCTTTT	TATAAAAGAA	CGTAACTTCA	TGTTAACGTT	21360
	TGACTTATAA	AAATGGTGGA	AACATAGATT	AAGTTATTAA	GGGCGCACGG	TGGATGCCTT	21420
	GGCACTAGAA	GCCGATGAAG	GACGTTACTA	ACGACGATAT	GCTTTGGGGA	GCTGTAAGTA	21480
50	AGCTTTGATC	CAGAGATTTT	CGAATGGGGA	AACCCAGCAT	GAGTTATGTC	ATGTTATCGA	21540

55

	GAGGAAGAGA AAGAAAATTC GATTCCCTTA GTAGCGGCGA GCGAAACGGG AAGAGCCCAA	21660
	ACCAACAAGC TTGCTTGTG GGGTTGTAGG AACTCTATA CGGAGTTACA AAGGACGACA	21720
5	TTAGACGAAT CATCTGGAAA GATGAATCAA AGAAGGTAAT AATCCTGTAG TCGAAAATGT	21780
	TGTCTCTCTT GAGTGGATCC TGAGTACGAC GGAGCACGTG AAATTCCGTC GGAATCTGGG	21840
	AGGACCATCT CCTAAGGCTA AATACTCTCT AGTGACCGAT AGTGAACCAG TACCGTGAGG	21900
10	GAAAGGTGAA AAGCACCCCG GAAGGGGAGT GAAATAGAAC CTGAAACCGT GTGCTTACAA	21960
	GTAGTCAGAG CCCGTTAATG GGTGATGGCG TGCCTTTTGT AGAATGAACC GCGGAGTTAC	22020
	GATTTGATGC AAGGTTAAGC AGTAAATGTG GAGCCGTAGC GAAAGCGAGT CTGAATAGGG	22080
15	CGTTTAGTAT TTGGTCGTAG ACCCGAAACC AGGTGATCTA CCCTTGGTCA GGTGAAGTT	22140
	CAGGTAACAC TGAATGGAGG ACCGAACCGA CTTACGTTGA AAAGTGAGCG GATGAACTGA	22200
	GGGTAGCGGA GAAATTCCAA TCGAACCTGG AGATAGCTGG TTCTCTCCGA AATAGCTTTA	22260
20	GGGCTAGCCT CAAGTGATGA TTATTGGAGG TAGAGCACTG TTTGGACGAG GGGCCCCCT	22320
	CGGGTTACCG AATTCAGACA AACTCCGAAT GCCAATTAAT TTAACCTGGG AGTCAGAACA	22380
	TGGGTGATAA GGTCCGTGTT CGAAAGGGAA ACAGCCCAGA CCACCAGCTA AGGTCCCCAA	22440
25	ATATATGTTA AGTGGAAGG GATGTGGCGT TGCCCAGACA ACTAGGATGT TGGCTTAGAA	22500
	GCAGCCATCA TTTAAAGAGT GCCTAATAGC TCACTAGTCG AGTGACACTG CGCCGAAAAT	22560
30	GTACCGGGGC TAAACATATT ACCGAAGCTG TGGATTGTCC TTTGGaCAAT GGTAGGAGAG	22620
	CGTTCTAAGG GCGTTGAAGC ATGATCGTAA GGACATGTGG AGCGCTTAGA AGTGAGAATG	22680
	CCGGTGTGAG TAGCGAAAGA CGGGTGAGAA TCCCGTCCAC CGATTGACTA AGGTTTCCAG	22740
35	AGGAAGGCTC GTCCGCTCTG GGTAGTCGG GTCCTAAGCT GAGGCCGACA GcGTAGGCGA	22800
	TGGAFAACAG GTTGATATTC CTGTACCACC TATAATCGTT TTAATCGATG GGGGGACGCA	22860
	tAGGATAGGC GAAGcGTGcG ATTGGATTGC ACGTCTAAGC AGTAAGGCTG AGTATTAGGC	22920
40	AAATCCGGTA CTCGTTAAGG CTGAGCTGTG ATGGGGAGAA GACATTGTGT CTTGAGTCG	22980
	TTGATTTTAC ACTGCCGAGA AAAGCCTCTA GATAGAAAAT AGGTGCCCGT ACCGCAAACC	23040
	GACACAGGTA GTCAAGATGA GAATTCTAAG GTGAGCGAGC GAACTCTCGT TAAGGAACTC	23100
45	GGCAAAATGA CCCCCTAACT TCGGGAGAAG GGGTGCTCTT TAGGGTTAAC GCCCAGAAGA	23160
	GCCGCAGTGA ATAGGCCCAA GCGACTGTTT ATCAAAAACA CAGGTCTCTG CTAAACCGTA	23220
	AGGTGATGTA TagGGcTGAC GCCTGCCCCG TGCTGGAAGG TTAAGAGGAG TGGTTAGcTT	23280
50	CTGCGAAgCT ACGAATCGAA GCCCCAGTAA ACGGCGGCCG TAACTATAAC GGTCCTAAGG	23340

EP 0 786 519 A2

	TGTCTCAACG	AGAGACTCGG	TGAAATCATA	GTACCTGTGA	AGATGCAGGT	TACCCGCGAC	23460
	AGGACGGAAA	GACCCCGTGG	AGCTTTACTG	TAGCCTGATA	TTGAAATTCG	GCACAGCTTG	23520
5	TACAGGATAG	GTAGGAGCCT	TTGAAACGTG	AGCGCTAGCT	TACGTGGAGG	CGCTGGTGGG	23580
	ATACTACCCCT	AGCTGTGTTG	GCTTTCTAAC	CCGCACCACT	TATCGTGGTG	GGAGACAGTG	23640
	TCAGGCGGGC	AGTTTGACTG	GGGCGGTCGC	CTCCTAAAAG	GTAACGGAGG	CGCTCAAAGG	23700
10	TTCCCTCAGA	ATGGTTGGAA	ATCATTTCATA	GAGTGTAAG	GCATAAGGGA	GCTTGACTGC	23760
	GAGACCTACA	AGTCGAGCAG	GGTCGAAAGA	CGGACTTAGT	GATCCGGTGG	TTCCGCATGG	23820
	AAGGGCCATC	GCTCAACGGA	TAAAAGCTAC	CCCGGGGATA	ACAGGCTTAT	CTCCCCCAAG	23880
15	AGTTCACATC	GACGGGGAGG	TTTGGCACCT	CGATGTCGGC	TCATCGCATC	CTGGGGCTGT	23940
	AGTCGGTCCC	AAGGGTTGGG	CTGTTCCGCC	ATTAAAGCGG	TACGCGAGCT	GGGTTTCAGAA	24000
	CGTCGTGAGA	CAGTTCGGTC	CCTATCCGTC	GTGGGCGTAG	GAAATTTGAG	AGGAGCTGTC	24060
20	CTTAGTACGA	GAGGACCGGG	ATGGACATAC	CTCTGGTGTA	CCAGTTGTCT	TGCCAACGGC	24120
	ATAGCTGGGT	AGCTATGTGT	GGACGGGATA	AGTGCTGAAA	GCATCTAAGC	ATGAAGCCCC	24180
	CCTCAAGATG	AGATTTCCCA	ACTTCGGTTA	TAAGATCCCT	CAAAGATGAT	GAGGTTAATA	24240
25	GGTTCGAGGT	GGAAGCATGG	TGACATGTGG	AGCTGACGAA	TACTAATCGA	TCGAAGACTT	24300
	AATCAAAATA	AATGTTTTGC	GAAGCAAAAT	CACITTTACT	TACTATCTAG	TTTTGAATGT	24360
30	ATAAATTACA	TTCATATGTC	TGGTGACTAT	AGCAAGGAGG	TCACACCTGT	TCCCATGCCG	24420
	AACACAGAAG	TTAAGCTCCT	TAGCGTCGAT	GGTAGTcGAA	CTTACGTTCC	GCTAGAGTAG	24480
	AACGTTGCCA	GGCAAAAAAT	GGATGCGATG	AGCCGCATTG	AGACCGCAAG	GTCTCTTTTT	24540
35	TTTATGTCTA	AAACGTCAAA	ATAAAAAGCA	AACACAAAGA	AAAATGGCTT	GGCGAAGTGA	24600
	AAACGTTTGA	ATCTGACGAA	ACGAGAAAAG	ArcGCAACGA	GTTTAGTAGA	GCTAAATGAG	24660
	TAAGyGAGAG	CCGAAGrAGA	GGAAAGAAGC	AAGCGATTGT	CACAAGTCAA	GAAAGGTTCT	24720
40	TAGCGAsGAT	GGTAGCCAAC	TTACGTTCCG	CTAGAGTAGA	ACTGGAAATG	ATAATTTAAT	24780
	AATGTACACT	TTCGATTGTC	TAAGTATGTA	CAACTTTAAT	TTTGTGTTTA	TATAAATTTA	24840
	AAATGATATC	ATCGAAAACA	AAATATTGTA	TAAATAGAGA	AGAGCAGTAA	GACGGTATCT	24900
45	AATTGAAAAT	GATCTTACTG	CTCTTTTATA	TACTTTATTG	AAATACAAAA	AGGAAATTA	24960
	TTATTATACA	ATAGACAAGC	TATTGCATAA	GTAACACTAA	CTTTTATCAA	AGAAGTGTTA	25020
	CTTTATAATT	AATGATTTTA	TTAGAGCGTC	TACATGCGGT	TTTAAAGCAT	CATCGTCTAT	25080
50	ACCGCCAAAG	CCTAATATAA	ATTTAGGGGT	TTTCTTATAG	TCTTGATCAT	CATCAAAATT	25140

55

	TCCATTTTTT	ACTGTAATTG	TAAAATGCAT	ACCCGTTTCA	GCACCTTGAA	TATCAAGCTG	25260
	CTCTTTGTAA	GGTTTCAATC	TTTTTAAAAT	ATAGGTTAGT	TTTCTACGAT	AAATTGCTCT	25320
5	CATTTTATTT	AAATGCCTTT	CAAAACCACC	GGAAGATATA	AACGTTGCAA	TAAGGTTTTG	25380
	CATATGAACA	GGTACAGTGT	TGCCTTCAAT	GTGATTTTGA	GAATGATATT	TTTTCATAT	25440
	AGAATAGGGT	AACACCATAT	ATGCAACTCG	ACAGCTAGGA	AAAATAGACT	TTGAAAATGT	25500
10	ACTGATATAA	ATCACTTTTT	CTCCTCTTGA	ATATAGACCT	TGAATTGCTG	GAATGGGTTT	25560
	GCCGAAATAT	CTAAACTCGG	AATCATAATC	ATCTTCTATA	ATAAATCGTT	CTTCTTTTTC	25620
	TTGAGCCCAT	TGTATTAATT	GAGTTCGTTT	TTTTAAGTCC	ATCACATATC	CAGTTGGAAA	25680
15	TTGATGGGAA	GGCGTTATAT	ATACTATATT	TTTTTGTGAT	TTAATAACTT	CATCTACGTT	25740
	TATTCATTA	TCTTCAACTT	CAATTTGTTC	ATATTCAACT	TGTTTTTTAT	CTAAAATATT	25800
	TTTGATTGGT	GGATAACTAG	GTTTTTCGAT	AATAAATGTT	GAAGTATAAA	GTAAATCGAC	25860
20	TAATTGATTT	ACTAATTGTT	CGGTAGATGA	GCCAATTATA	ATTTGATTAG	GATCACAAAT	25920
	TACGCCACGA	TTAGTAAATA	AATAAAATGC	CAGTTGAAAC	CGCAAATGTA	ATTCTCCTTG	25980
	AAAATGTCCT	CTACGTAATT	GATTTAAATG	ATTTGTATCA	TAAAGATCTT	TGGAATACTT	26040
25	TCTGAAAAGT	TCTATAGGGA	AATGTTTCGT	ATCTATTTCA	TCCAAATTAA	AAGCATAATC	26100
	ATAAGCTTCA	TCACTCGCTT	TTGGTTTATA	TGAATCATCA	TCAAAAAGAG	AGGGGATAGG	26160
30	TTGATTGTTT	AAAATTGTTA	AAGATTCAAT	TTCGGACACA	AAATATCCAG	AGCGAGGTCT	26220
	TGAATAAATG	TAACCTTCGT	CTAATAGAAG	TTGATATGCA	TGCTCTACGG	TTGTTTGGCT	26280
	AATAGATAAA	TGTTTGCTTA	ATTGTCTTTT	AGAATAAAAT	TTATCGCCTT	CTTTAAATTG	26340
35	ACCTTCAATT	ATTTGTTTTT	TTAATTTTTT	ATAAAGTTGA	TGGTATAAAG	TGTTTTTCAA	26400
	TTTTATAACT	GACCTCCTAA	ATTTATCTTA	TTTTGTACCT	TTTTAAATAT	CAGTTTATAC	26460
	ATTACAATGT	ATTTAATCAA	CTTGAAAAGG	GGTTTTATGT	ATAATGAGTA	AAATTATTGG	26520
40	ATCAGACAGA	GTCAAAAGAG	GTATGGCTGA	AATGCAAAAA	GGCGGCGTTA	TTATGGATGT	26580
	CGTTAATGCT	GAGCAAGCAA	GAATTGCAGA	AGAAGCTGGC	GCGGTAGCAG	TTATGGCATT	26640
	AGAACGAGTA	CCTTCTGATA	TTAGAGCTGC	TGGTGGTGTT	GCACGTATGG	CAAACCCATA	26700
45	AATTGTAGAA	GAAGTAATGA	ATGCTGTTTC	TATTCCAGTC	ATGGCTAAAG	CACGTATTGG	26760
	TCATATCACT	GAAGCAAGAG	TATTAGAGGC	GATGGGTGTT	GAATATATTG	ATGAATCAGA	26820
	AGTGTTAACA	CCAGCAGATG	AGGAATATCA	CTTAAGAAAA	GATCAATTTA	CAGTACCATT	26880
50	TGTATGTGGA	TGTCGTAATT	TAGGTGAAGm	TGCGCGTAGA	ATTGGTGAAG	GTGCTGCTAT	26940

55

	ACAAGTTAAT TCAGAAGTTA GTCGATTGAC TGTAATGAAT GATGATGAGA TTATGACTTT	27060
	TGCGAAAAGAT ATCGGTGCGC CTTATGAAAT TTTAAACAA ATTAAAGACA ATGGTCGTTT	27120
5	ACCGGTAGTT AACTTTGCAG CTGGTGGCGT TGCGACTCCT CAAGATGCTG CTTTAATGAT	27180
	GGAATTAGGT GCTGACGGTG TATTCGTTGG ATCAGGTATT TTTAAATCAG AAGATCCAGA	27240
	AAAATTTGCT AAAGCAATTG TTCAAGCAAC AACACATTAC CAAGACTATG AACTAATTGG	27300
10	AAGATTAGCA AGTGAACCTG GCACTGCTAT GAAAGGTTTA GATATCAATC AATTATCATT	27360
	AGAAGAACGT ATGCAAGAGC GTGGTTGGTA AGATATGAAA ATAGGTGTAT TAGCATTACA	27420
	AGGTGCAGTA CGTGAACATA TTAGACATAT TGAATTAAGT GGTCATGAAG GTATTGCAGT	27480
15	TAAAAAAGTT GAACAATTAG AAGAAATCGA GGGCTTAATA TTACCTGGTG GCGAGTCTAC	27540
	AACGTTACGT CGATTAATGA ATTTATATGG ATTTAAAGAG GCTTTACAAA ATTCAACTTT	27600
20	ACCTATGTTT GGTACATGCG CAGGATTAAT AGTTCTAGCG CAAGATATAG TTGGTGAAGA	27660
	AGGATACCTT AACAAAGTTGA ATATTACTGT ACAACGAAAC TCATTCCGTA GACAAGTTGA	27720
	CAGCTTTGAA ACAGAATTAG ATATTAAAGG TATCGCTACA GATATTGAAG GTGTCTTTAT	27780
25	AAGAGCCCCA CATATTGAAA AAGTAGGTCA AGGCGTAGAT ATCCTATGTA AGGTTAATGA	27840
	GAAAATTGTA GCTGTTTACG AAGGTAAATA TTTAGGCGTA TCATTCCATC CTGAATTAAAC	27900
	AGATGACTAT AGAGTAACTG ATTACTTTAT TAATCATATT GTAAAAAAG CATAGCTTAA	27960
30	TGTATGCTAA ATCAACGAAT TATTGATATT TATAGATTG TTGAGAAGAA AATATCTCCT	28020
	TCAAACCTAG CTTTGAGGA GTTATTTTTT ATGTCAAAAT TAAAAATGAT AAAAAATAAA	28080
	GCTATACATA AGAAAAAAC CCTTCAAAGA GACTGAGAAT AGTCAAAATT TTGAAGGGGT	28140
35	TAATTCGATG TTGATGTATT TGTAAATAA AGAATCCAGC GATTGCAGCT GAAATGAAAG	28200
	ATACTAGTGT TGCACCGAAT AATAATTTCA AACCAGCG GGCAACTGTA TCTCCTTTTT	28260
40	TGTCATTAAG TGATTTAATC GCACCTGAAA TAATACCGAT AGAGCTAAAG TTAGCAAATG	28320
	ATACTAAGAA TACAGATGTA ACACCTTTTG CGTGTTCAGA TAAATCACTA AGTTTACCAA	28380
	GTGCTTGCAT TGCTACAAAT TCGTTAGATA ATAGTTTTGT CGCCATAACT GAACCGGCTT	28440
45	GAACTGCATC TTGCCATGGC ACACCGACTA AGAATGCAAA TGGTGCAAAG ACAAACCAA	28500
	TTAATGTTTG GAAATCCCAA GAAATAGCGC CACCTGAAAC TGTACTAAAG ATATTGCTTA	28560
	CAATTCATT TAATAGAGCG ATAATGGCAA TGTATCCGAT TAACATTGCG CCTACAATGA	28620
50	CAGCTACTTT AAATCCATCT AAAATATATT CTCCTAGCAT TTCGAAGAAT GATTGTTGTC	28680
	TTTCTTCAGT TTCTTCAACT AATAATTTGT CATCTTCTTC ATTAACTTTA TAAGGGTTAA	28740

55

	TAGGTTCAAT TAAGGTAAAG TATGCACCGA TAATTGAAGC AGAAACAGTC GACATTGCTG	28860
	AAGCTGTTAA TGTGTATAAA CGTTGCTTAG GTATGTATGG TAATTGTTTT TTAATTGAAA	28920
5	TAAATACTTC AGATTGTCCC AAAATTGCTG CAGCAACTGC ATTGTATGAT TCTAAACGTC	28980
	CCATACCATT AATTTTAGAA ATTAAGAATC CTAAAACATT AATGATTAAA GGTAAAATCT	29040
	TTGTGTATTG AAGGATACCG ATAATCGCTG AAATAAATAC GATAGGTAAT AATACACTGA	29100
10	AGAAGAATGG TGGTTGCTTA GGATCGATAT ATTGAATACC ACCGAATACA AAGTTAACAC	29160
	CATCTGCTGC TTTTAATAAT AAGTAGTTAA AACCGTTTGA AATACCACCA ATAACCTTGA	29220
	TTCCCATTTGT AGTTTTAAGC AAGATAAATG CAAAGATAAG CTGAATTGCA AGTAAAATTC	29280
15	CTACATATTT CCAGCGAATA TTTTTCCTGT CTGAGCTAAA TAGAAACGCA AGTGCTAAAA	29340
	AGAAGATAAT TCCGATAATC CCAATTAGAA TATGCATATA TTTCTCATT CTTTAGTTTT	29400
20	TTCTACaAtC TATCATACAA TAAAATGGAA GGGCTAACAT CATAAATTTT TGAAAATATA	29460
	AAAACAAATT AATTGAAAAA GGTCAAAATA GGTCAATATA TATAGTCAAA GAAGGTCAAA	29520
	AAGGGGTGAT ATACATGCAC AATATGTCTG ACATCATAGA ACAATAaTCA AACGTTTATT	29580
25	TGAAGAGTCG AATGAAGATG TCGTTGAAAT TCAGAGAGCG AATATCGCAC AGCGTTTGA	29640
	TTGCGTACCA TCACAATTAA ATTATGTAAT CAAAACACGA TTCACTAATG AACATGGTTA	29700
	TGAAATCGAA AGTAAACGTG GTGGTGGTGG TTACATCCGA ATCACTAAAA TTGAAAATAA	29760
30	AGATGCAACA GGTATATATTA ATCATTTGCT TCAGCTGATT GGACCTTCTA TTTCTCAACA	29820
	ACAAGCTTAT TATATTATTG ATGGGCTTTT AGATAAAATG TTAATAAATG AACGTGAAGC	29880
	TAAAATGATT CAAGCAGTTA TTGATAGAGA AACGCTATCA ATGGATATGG TTTCTAGAGA	29940
35	TATTATTAGA GCAAATATTT TAAAACGTTT GTTACCAGTT ATAAATTATT ACTAAATGAA	30000
	ATGAGGTGTT GAAGTGCTTT GTGAAAATTG TCAACTTAAT GAAGCGGAAT TAAAAGTTAA	30060
40	AGTTACAAGT AAAAATAAAA CAGAAGAAAA AATGGTGTGT CAACTTGTG CTGAGGGGCA	30120
	CCATCCGTGG AATCAAGCTA ATGAACAACC TGAaTATCAA GAACATCAAG ATAATTTGCA	30180
	AGAAGCATTT GTTGTTAAGC AAATTTTACA ACATTAGCT ACGAAACATG GAATTAATTT	30240
45	TCAAGA	30246

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 14333 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

	TATTCCTCCCA TCGGTTTATT AAATCGTCCA TTTCAATACT GTTTTTCCTCC AAGATGTCGA	60
5	TAAATCCATT TCAAACGCTT GGACGATATC TTGCATCGTA CATACTTAA TTTTATGTCC	120
	TTTAAATAAT GCTAACTTTT CAACTATGTC TGGGTACTTA CGATATAAAT CAACAACCTG	180
	CTCAAAATCT TTAGAGCCGC TTCGACTACT ACCAATCAAC GTTAATCCTT TTTCAAGTAC	240
10	TAATCGTGTA TTCACTTCCA CGGGTAATTC ACTTACGCCT AACAAAGCAA TACTGCCTTC	300
	TGGTGAAATA TGTTCACATA TTTGTTGAAG TGCAACTTGA CTTCTTTTAC CTCCAACACA	360
	TTCAATGCA TGATCAATTT TAAGATCATC TGGTATTTGA TTTACTGTAA AGATGTCATC	420
15	TACAAATGAA AAATGACTTA ATTTATAGTC TGTCTTACCA AATACATAAG TTTTAGCTTC	480
	TGGGTACAAC TTACGTAGCA AAATAGCAGT AATATAACCT AAGTTACCAT CACCCCAAAT	540
20	ACCAAAGCTG GTTTTCAAAG GTATAGATTT ACGTTCAAAT CGTTGTATAG CATGATAACT	600
	TACTGACACT AACTCTGTGT ATGAAATCGT ACTCAAATCA ATGTCATTAG GCAGCGGAAC	660
	GATACGATCA TGTGCCATCA CAACGTAGTC TTGCATAAAA CCATCATAAC CACTAGATCT	720
25	AAAATAACTA GAGGCTAAGT AATTCTCCGC AATAATATGA TGTGCTCTG TAGGTGTATT	780
	CGGTACCATT ACTACTTTCG TACCTTTTTC AAATACCCCT TTACTATCAA ATACAACCTC	840
	ACCAACAGCT TCATGAACTA ATGACATTGG TAATTTTTTG CGTAGTACAT TTTTCTCTCT	900
30	TCGACCTGTG TAATACCTTT GATCAGCTGC ACAAATAGAC AAGTATAAAG GTCTTACGAT	960
	GACATGATTA CCATAAATAT CAACATTATT ATATGTGACG TCGAACTGTC TCGGTGCAAC	1020
	GAGTTGATAT ACTTGATTAA TCATCGGCAA TATCACCTTG AATAATGGCA TTTGCTACTT	1080
35	TTAAATCATA CGGTGTTGTC ACTTTAATGT TGTATAGTTC TCCaCGTACC AATTAACTG	1140
	CATGTCCAGA TTCGACAATG ATTTTACATG CATCTGATAA GATTCTTTT TGTTCACTAC	1200
40	TTAAGGCGCG ATAACATCTT TGTAATAATT TAATATTAAA TGATTGTGGT GTTTGGCCTT	1260
	GATACATTTC ATTCCTTACA GGGATACTGT GTATGTTCTG TTTATCTTTA GACATTACAA	1320
	TCGTATCAAT TGCTTCAATG ACTGTATCTA CTGCACCATA TTTTGCTGCT ACTTCAATGT	1380
45	TCTCTTTAAT AATACGTTGA GTTAAAAATG GTCTTACGGC ATCATGAGTT ACAATCACAT	1440
	CATCATTATT AATTCCATTT ACATTGCGAA TATGGTCGAT AATGTTTATA ATTGTTTCAT	1500
	TTCGATCCGT ACCACCTGCA ACTACTTTGA CACGTTGATC TGTAATGTTA TATTTTTTTA	1560
50	AAATATCCTG TGTATGGGAA ATCCACTGTG CTGGCGTTGC GATAATAATC TCATTAAATT	1620
	CACTCACTAA AATGAACCTC TCAATTGTAT GGATTAAAAT CGGTTTATTA TCAATATCTA	1680

55

	CTGCATAAAT CATGTTGTCC TCCATTCTGT CATTACATCA TTTCCATTTA TACATTACTG	1800
	ACCTATGCCC GCACATAAGC CTAACCTATT GCTCACTTGC CTCTTTTATT AATCCAAAGA	1860
5	TAGTTGTCAC AATAGTGTGA TAATTTTTTA TAAAAATGTA TTTTGTAAAC TGACCATTCT	1920
	AAGTTGTTTT GCCATGCAGT TAATCATTAA CTCTGACGAT ATTAAATTGT TAAAGGTATT	1980
10	AATGTTTACT CTTTTTCAAA TTCATTATTA CTGCCATCAT TTTACCATAT ATTATAATAA	2040
	ATTTATCTTA TTAAGTGGCT GTACTTGATT TTCACTTTAA AAATTATCAA ATATTGCCAT	2100
	CTCATTTTAA GTATACAAAA TGCAAAACAA CCGATTCAAC AGCATATTTT ACACAAGTAA	2160
15	ACCGGCTATT TATCAACGTA TATTCGAAGA TGAATTATTT CGATAGTATC TATAGACCAG	2220
	ACGGCATTCT CACTTTCATA GCTATAACTA TACCAGCGTT TTCGTCCTCA AAGGTGCATA	2280
	CTAATAAATC GTAAACATGA CTTTATCAAA TCGTCTTTT TGTTAACTA ATTTATCAAA	2340
20	TGTCTCCGGG CCTTTTTCTA ACGGTAAAAA ATGAGAAATA ATAGGCTTTA CATTAATATC	2400
	TTTCGTCTTC ATATAATGTA AGGTTGCCGT CCACTCTTTG CCCGAAAAT TACTGGACAA	2460
	ACAGTTCCAA GAGCCACATA CTGTCAACTC GTTACGCAGA ATTTTTTCAA AATGAACGCG	2520
25	ATCAATCTCA ATATCATCAT ATGGTATTCC GAGTAATACC ACCTCGCCAC CTTTTTTAGG	2580
	TAGCGTCAAT ATTTGACCAA TCGTAACTTT AGCACCTGAT GATTCTATAG CTAAATCGAT	2640
	TTGATTGGCG TAATGATTTT CGATGAATTT CTCAAGATTT TCTTCTTTTG AATTGATTGT	2700
30	TTGATGTGCG CCCAATGATG TTGCAATATC TAGTTTATGC GCATCTATAT CTATAGCGAT	2760
	GATATGTGCA GCACCAAATA TTCGTGCCCA TTGAATAGCT AACAAACCTA TACTGCCACA	2820
35	CCCCATTACT GCAACAGTCA TACCAGGTTG TATATTGAT TTATAAAACC CATGCGCAAC	2880
	AACGGCTGAT GGCTCAACCA TTGCTGCTTC AATGTAATCA ACATTGTCTG GAACCTTTAA	2940
	AACATTTTGC GCTGGCAATT TGACATATTC CGCGAACGAT CCAGGTTTAT ATGAGCCAAT	3000
40	GACGAATAAC TTTTCACATC GTGCATATTC ACCTTTTAAA CAATACTCGC ATTGATAACA	3060
	AGGTATTGCT GGGCAACCTG TCACTTTGTC GCCCACATTA ACATGCGTAA CATCACTTCC	3120
	AATGGCATCT ACTACACCTG AAAATTTCATG ACCAAATGGC ATACCTTTAA TGTATGGCCC	3180
45	CATTTTTTTG TATCGTGACG TGTCTGAACC ACATATGCCA GTCGCTCGTA CTTTAATAAT	3240
	AACGTCATTG GCACTTTCAA TGAATGGCTT TTCATTATCC TCATACCGTA AATCTTCCAC	3300
	GCCATATAAT TTCAATGCTT TCACTTGTA ATCACCTCAA ATTTGATTTA ATTCACAAC	3360
50	TTTTCTTTT TAAAAATACC TGTCGCAAAA TAACCTGCAA TGACAATGGA ATTACTTACG	3420
	AGTAAATGTT CCATATAAAA ATCAGTGATT TGTCTTAATG GCCCAAGCAT AAAAGTTAGC	3480

55

	TGCTTTAATA CCTTCGCCGG ATTTTAAATG TTGATACGCC TCGTCCCATT TCGAAATATC	3600
	ATATATTTTT GTCACCAAAG CTTCAGCATT TACTAAACCA TCCGCCATAA GTTGCAATGA	3660
5	AGGTTCCCAA TCTGCTGGCT TTTGACTTCT ACTACCAACA ACTGTTATTT CTTTTGAAT	3720
	CACTTTTTCC ATATCAAATG GAATTCAGC ATCCTTAAAA ATACCTATTT GACTGTAGAA	3780
	ACCTTTTTTG CGTAAAATAT CCAAACCTTG TCGTGCTGCT GGAAGTGCAC CTGAACATTC	3840
10	AACAACAACA TCTGCACCGT AACCGTCTGT AATTCATTG ATATACGTTT TTAAGTCTGT	3900
	TTGTTGTAAA TTGACTACAT AATCCATGTG CAATGCTTCT GCTTTATCTA ATCTGACTTT	3960
	GTCATTGTCC AATCCAGTTA CCACAACAGT TCGCCTTTA CTTTTTAACA CTTGTGCTAC	4020
15	AAGTAATCCG ATTGGCCCAG GTCCCATTAC AACTGCTACA TCGCCTGAAT TGACTTGAAT	4080
	CTTAGAAACG CCATGATGTG CACATGCTAA TGGTTCTGTC ATAGCTGCAG ACTGATACGA	4140
20	TATTCGCTCTG GAATATGATG CAACTTTTCT TCACGTGCAA TGACATAATT AGTAAATGCG	4200
	CCATCAACTT GTGTTCGAAT ACCTTTTCGA TGGTTGCATA AATTATAGTC TTTTGATTTA	4260
	CAGTATTCAC ACTCATTACA AACATAGAAT GTCGTTTCAG aTGtGACACG GTCACCAACT	4320
25	TTAAATCTT TAACGTCTGC TCCAACCTCA ACGATTCAC CAGAAAATTC ATGACCTAAT	4380
	GTCAGTGAA AATTAACTTT ATAATGACCT TCATAAGTAT GAATATCTGT GCCACAAATT	4440
	CCTGCATAAT GTACTTTAAT CTTTACTTTA TCATCTAGCG GTGTTGCAAC TTCTTTATCA	4500
30	AGAAGTTCTA AGTGGCCATG TCCTTCTCTT GTTTTTACTA AAGCTTTCAC CACAAACACC	4560
	TCGATTTTTA ATTGAATAGA CTAAATAGTT TAAAGATAAG ATAGTTAACG ATATTACCAC	4620
	CTTGATCAAT ACTTGAAATT TCAGATGAAC CTTTTGGCAT TTGTACATTC GTACCTTTCG	4680
35	CCATATCTGT GAAAATGGGT GCTACGTCTG TTGCAATATA TAGTGAAATT GCAATCATAA	4740
	TCGTACCCAC AATGACAGAA TGAATAATGT TTCCTCTTGC TGCACCAACA ATAAACGCGA	4800
40	CAACAAATGG TATCGTTGCT AAGTCACCAA AAGGTAGTAC TTGGTTTCCT GGTAAAATAA	4860
	CGGCTAATAA AACAGTGATA GGTACTAAAA TTAATGCTGT CGAAATAACT GCTGGATGAC	4920
	CTAATGCTAC AGCCGCATCC AATCCAATAT AAATTTACAG TTCGCCAAAA CGTTTATTTA	4980
45	GCCATGTTCT TGCAGACTCT GAAACTGGCA TTAAACCTTC CATTAAGATT TTTACCATTTC	5040
	TAGGCATTAA TACCATTACT GCAGCCATTG ACATTCCTAA ATTAATGATG TCTCCAGGTT	5100
	TGTAACCTGC TAACACACCA ATACCTAAAC CTAAAATTAA GCCGACAAAT ATAGACTCTC	5160
50	CAAATGCGCC AAAACGTTTT TGAATTGTTT CAGGATCAGC ATCTAACTTA TTCAGACCGG	5220
	GTACTTTTTG TAACAATTTA ACTAAGTAAA TACCTGGTGC ATAAGAAATT GTACTTCCTG	5280

55

	CTACTTTCAA ACAGATAATT TGGAAAATAA CTGCTGCTAA TAACGCTTGC CAAATACTGC	5400
	CTGATACGGC ATAAACCATT GCTGCTGTAA ACGTATAATG CCAAAAATTC CAAATATCTA	5460
5	CATTCATCGT CTTTGTCACT TTAGTTACTA GCAATACAAC GTTAACTATG ATTCCGAGTG	5520
	GAATAATAAA TGCTGCGACA GATGATGCCC AAGCGATAGA TGATGTTGCT GGCCAACCTA	5580
10	CATCAATCAC ATTCAGACTG ACGCCTAAAT TTTTAACCAT CGCTTGTGCT GCTGGCCCTA	5640
	AATTTTAAAC TAATAAATCG ATGACTAAGA AAATCCCTAC AAAAGCCACA CCTATTGTTA	5700
	AACCAGACCT AAATGCCGCT CCAATTTTCT GCCTAAAGAA TAGGCCAAGC AAGAATATGA	5760
15	CAACCGGTAA AATAACAGTt GCACCTAAAT CTAAAAATCC CCTTACAAA TCAGTGAAGT	5820
	AACTCATATT TAAACCCTCC CTGTTATATA TGCATTGTCA CGATACTTTC CGATTGTGAT	5880
	TACATTTGAC GTTACAGTCA TTTCAACGAC AACCCTTGCT AAATTCGACT GCAGTCCTTT	5940
20	TGAATTACAG tCACTGCGTT TCTATGTCAT CAACAATCAT TTGTCGTGAT AGTCATTTAT	6000
	ATGCAATTTG CATATATTAA TATGTTATCG ACCCACGTTA CATATCAATT CCGTTATTTT	6060
	TGTAACCTCG TTAAGATTG TTGTTTTGTT TCTTCAATAC CAATACCAGT TAAGAAATTA	6120
25	CGTGCGTTGA TAACTGGGAA TTTATATTCT TTTTTGTGCA TTGCAGTTGT AACTAATAAA	6180
	TCTGCAGTGT CTTCATAAGG TCCAACCTCT GTAATTTTGA TTTGTTTAAT ATCTACTTTA	6240
	ATATTGTGTT CCTTTGCCAT TTCTTCAATT GCAATTATTA CTACTGTTGA CGTTGCAATA	6300
30	CCTGCACCAC ACGCTACTAA TACTTGTTTC ATTTTCAATT CCTCCAATTA ATTTTITAGTT	6360
	ATATTCCAAA TAATCATTGA TTAGTGTGTC TAAATTTGTT TCATCTTTTCG TTCGTAGAAT	6420
35	CTGCTCCAAT TTTTCTTCAC TTTGAAAAAT TTGCATCAAC TGTGTGAACA GCTTAAGTTG	6480
	ATCATCTACT TTATCCATTG CTAACATAAA AACGATTTTC ACTTCTGTCT GTTGATCAAG	6540
	TGTTCCCAT TCAATAAACG GCACCTCTTT TTCTAGAACA GCCACACCTA TCGTTCTATG	6600
40	GTTAATATGT TCGACATCTG TATGCGGTAT AGCGACCGAA CATAGATGCG TTGGTAAACC	6660
	AGTAGCAAAT TCTTTTTCTC TGTCGATGAC TGCATCTTTA AACGTTGACT TCACGAACCC	6720
	ATTTTGAAAT AACACATCTG ACATTTGTGA CAATACGGAT TCTTTATCAG TTGCCGACAA	6780
45	ATTGAGCATT ATATTTTCTT TATGCACTAA TTGCTGTCCC ATCCATTTTC CCTCGCTTCT	6840
	TTATTTGAAT AATTTTTTAA AATCTCATTT ACATCAGAAT TTTTGCGACT TTGTATGATG	6900
	CGCTTAATTG CGTCATTGTC TTGCGCCACA TCTCTCAATT GTAGTAACGC TCTTAAGTGT	6960
50	GTCACTTTAT CAACAGCAGC AATAGGTACA ATAATATGGA TTGCTGTGCC ATCTGACATG	7020
	TATATTGGTT CTTGTAATAT CAACATACTC ATCGCTGTTT TATGTACATG CTTTTCAGAG	7080

55

	TGCATCTCAT GAATATATTT AATATCAATA AAATGATTAG CAACTAACAC ATCACTTGCT	7200
5	TTAGCAATAG CTTTCATCAAT ATTTTCAACA TGATGCATTG TTTTCACGTG CCTTGCCGGT	7260
	ATCAAGTCAG CTAAATCTAA TGyCTwATTT tGTGtGACaA TCGATCCATT AATGGTTGAA	7320
	ATTGAATTAT AATTGGCAAT AAAATCTTCT AAACCATCAC GTAGTcTGTA ATGTCATTAA	7380
10	CTGTCGTTGT GCGTTCAATT AATGCCATTA ACTTGTTTAT TTCCTTATCA ATGTCAGCCG	7440
	ATTCTTATT AATGTACTTC ATCACTTCTT TACGTAACCT TCGTTGCTCA TTTTCAGATA	7500
	AAGCTACTTT TGTGATAAAT AATTTTTTAT GTGTTAGGAC AAACATTGGT GAAAAGACGA	7560
15	TGTCATAATC TAATGTGTAA TTTTCAAATG TTCTAAGTGA AATCGCATCT AAGAAAATAA	7620
	TTTCTGGAAA TAAGTTTCGC AACTCGTATA ACATCATTTG TGATACTGAC GTGCCTTGTC	7680
	TACACACGAT AATAGCTTTT ATCTTGCCAT CGAAGTTTTT ATCTTGACGT CTCAAACTAC	7740
20	CTCCGAACAA CATGGTTAAA TATGCTATTT CATTATCAGG CAACGATTTT CCGAAATATT	7800
	CAGTTAACGA TTGACATGAT TGTTCACCA TATGAAATAA GGATTGATAA TTTCTTGTA	7860
	AAGGATTAT TAATTCATCA CGATCCGTTA AGTTATATTT AATCCTATAA AAAGCAGGCG	7920
25	TTAAATGTAA CAAGAGTTGC TGTGATAATT TCTCCTTATC TTCAATGTTA ATAAAAGTGA	7980
	TTTGTTCAAA ATGGTGAATC ATTTGAGCGA TGGCCATCGT TAAATTCGAT ATGCTATCTG	8040
30	ATTCTTGCAA ATCAGTCCAT TGCACACTTG TTGAAAGTAA GTGTAATGTC AAATATAACT	8100
	TTTCCGCTTC TGGCAAATCC GGCTCATGTT GCGTCATAAT CTCCGTTGCT TGATATTCTT	8160
	TCGTATCCCT CAAATACTGA TAATTAATAT TTAATGGATT CATCACATGA CCACTTTGAA	8220
35	TTCGTCTACG AATCACACAA AGGACATAAG GCAATGAACT AAGTGATTG TCTATAAAGC	8280
	GACTCTTCAA AAATTGTTCT ACCTGTTTGA TCTTGCTTTT TTGATATGCG ATATCTTCGA	8340
	ATGTTAAGTT GAGCGCCTTT AAAACTTCAC TTTTAGTAAT ATCATGATTC AACCTTTGAT	8400
40	CAATCAACTT AATGAAGAAA CGGCGAACTT CAAATTCATC ACCAACAATT TCATAACCAT	8460
	GTTTTGAGA ATACTTAAGT GACAAACCAT GATTTTCCAA TTGCTCTTTC ACATGATTTA	8520
	TATCGTGAAT GACAGTATTT TTAAGTACTT GTAAATCAAT TGAAAAATGG TTTAGAGACA	8580
45	TTGCGTTTTT CTTACTAAAA AGCATGAGCA TTAAATAATA ACGACGTGTT TCTATGCTAA	8640
	AAATGACATT GTTGCCGTTT AACATTTGCT GCTCCGATAC ATCTCGCTTG AATAACGTCA	8700
	TGATTTTACA ACTTACAATA AAATTTCTT GGCTTGTTCT TTCAAGTTTT GGATAACCCT	8760
50	CTTGTTCAAG CCACAAATTG ATTTTTTGAA TGCATATCC TAGTTGTCTA CGAGACAAAC	8820
	CAAATATCGA TTCAAGTTCT TTACCATGAA TAGTAGGATT CAATACAATT TCTCTGAGTA	8880

55

	TCAATCGTCA CACCGATGTA CACACTTTGA ACACATATTT TCAAAATGAG CATGTACATC	9000
	ATTGTGATGT TTTAACAACA TTTCAATTAT ATCTATATTT TTTGTGATTT TAATCTTTTA	9060
5	AAATAAAGCA ATTGAAATTT TTGCATATAT TTTTGTGTTT TGTGTTTTTT TGAAGCATTT	9120
	TTAACATACA TATCTCAATC ATTATCAAAT TGTCATGACC ATTGTAACCC AATACAAAAA	9180
10	CCCTAAGGAC GCTTATATCA GGCGCCTTAG GGTAACTGT ATCTATTTAA TTAAGTATTA	9240
	TTATTTCGTAT GTACGTAAC TATGGTCTAT CAAGTTCCAC ACTTCTTCAA CATCAACTGC	9300
	TGTAGCAAAA TAAGCATTGG CAGGCTTACC TGTAACATGA TTTAAATCGA CAGCCATAGT	9360
15	GCCATAAGTT AGTGGACTTT GATGTTCAAT GTCGATATTA ACGGGTACCA TTGTAAACAA	9420
	TTCTGGTTGT AACAAATACA AAATTGTACA AGCATCATGT ATTGGACCAC CATCCATATT	9480
	AAAGTGAGTC TTGTATGTCT TCTTAAAGAA TTGCAATAAT TCTACGACGA ACTGTGCAAC	9540
20	AGGATTATTG ATACTTTCAA AGCGTTCAAT CACGTGATCG TCGGCTAAAA CTTGATGTGT	9600
	TACATCTAAA CCAAACACAT TTATAGTAAT CCCACTTTCA AAAACACGCT TCGCTGCTTC	9660
	AGCATCTACC CAAATATTGA ATTCTGCTGT AGGCGTCCAA TTTCCAAATG TACCACCACC	9720
25	CATCAAAGTA ATAGATTCAA TATGCTCAGC GATTCTTGGC TCACGAATCA ATGCCGTTGC	9780
	TACATTGTA AGAGGACCTG TCGCTACAAT TGTTACAGGT GTATCACTCG TCATCACTTT	9840
	GTTTATAATC ACATCTGATG CTGGCATTGC AACTGCTTGA CGTGATGGTG TCGACGGTAG	9900
30	TTTCGGACCA TCTAATCCAG ATTCCCCATG TATTTTCAGAA GCAAAGGCAG CTGGTTTAAAT	9960
	TAACGGCCTA TCCGCACCTT TCGCTACTGC TATATCTTGG CGTCCCATAA TATCCAATAC	10020
35	GTTCAAGGCG TTTGTGCTAT TCTTGTCAAC TGATTGATTA CCTGCGACTG TTGTTACAGC	10080
	TAATATCTCT AGTGGACTGT CAATTGCCCC CGCTAAAATT AATGCTATTG CATCATCGTG	10140
	TCCTGGATCA CAATCCATAA TAATCTTTCT TTTCAATTTAT ATATCCACCT TTCTTAAGTT	10200
40	GTTATCGATA GCTTATGTAT ATTTATTTAT GTGGTGAATC ATGTTTATTT TGAAAAATAG	10260
	TTTTAACTTT CTCATATTTT TGGATACAAA CACTATTTAT CTATTTTATG GCTTATAAAT	10320
	TTATCCGATA TGCCTTATCA ACCTACCTCG CTAAAAATAG GATGTCTACA TATCTATACC	10380
45	GACTTTTGTC AACTCATTTT CACAACAATA TAAACAGCAA TTTATATGAT TGTTACATGA	10440
	TTCAAACAAT TTTTATGAAA AATATTTTCA TACACAGAAT ATATATTGAT ATTAAATTTT	10500
	TCAAAAGCTA TATTGAGAAT AATTAGGAGG GATGTTGATG AAATCTTTAT TTGAAAAGC	10560
50	ACAGCAGTTC GGCAAGTCCT TTATGTTACC TATCGCAATC TTACCAGCTG CAGGTCTATT	10620
	GTTGGGTATC GGTGGTGCAT TAAGTAATCC AAACACCGTT AAAGCATACC CTATTTTAGA	10680

55

AAATTTACCG GTCATCTTTG CAATTGGTGT CGCAATCGGA TTATCTAGAA GCGATAAAGG 10800
 TACTGCAGGT tTAGctGCGC TGCTCGGTTT CTTAATTATG AACGCAACTA TGAATGGCTT 10860
 5 ATTAACATATC ACGGGCACAT TGGCAAAAGA TCAGCTTGCA CAAAATGGAC AAGGCATGGT 10920
 GCTCGGTATA CAAACGGTTG AAACCGGTGT TTTTGGCGGG ATTATCACAG GTATTATGAC 10980
 CGCAATACTT CACAACAAAT ATCACAAGT GGTATTACCA CCGTATTTAG GTTTCTTTGG 11040
 10 TGGCTCTAGA TTTGTCCCTA TTGTCACAGC ATTTGCCGCA ATCTTTTTAG GTGTATTGAT 11100
 GTTTTTTCATT TGGCCAAGCA TACAAGCCGG CATTTATCAT GTTGGTGGAT TTGTAACGAA 11160
 AACAGGTGCC ATCGGTACTT TTGTTTATGG CTTTCATCTTA AGATTGTTAG GTCCACTCGG 11220
 15 TTTACACCAT ATTTTTTACT TACCGTTTTG GCAGACGGCA CTTGGTGGTA CTTTAGAAGT 11280
 CAAAGGGCAC TTAGTCAAG GTACGAGAA CATCTTCTTT GCTCAACTTG GTGATCCAGA 11340
 20 TGTGACGAAG TATTATTCAG GTGTGTCACG CTTTATGTCA GGCCGTTTTA TTACGATGAT 11400
 GTTCGGCTTA TGTGGTGCCG CACTTGCAAT TTATCACACA GCTAAACCTG AACATAAAAA 11460
 AGTTGTCGGC GGTTTAATGT TATCCGCTGC ACTCACTTCA TTTTAAACAG GTATTACCGA 11520
 25 ACCTTTAGAG TTTAGTTTCT TGTGTGTCGC ACCTATTCTT TATGTAATCC ATGCCTTCTT 11580
 TGATGGATTA GCATTATGA TGGCAGACAT TTTCAACATT ACAATTGGTC AAACCTTCAG 11640
 TGGAGGCTTT ATCGATTCT TACTCTTTGG TGTGCTACAA GGTAATAGTA AAACAACTA 11700
 30 CCTATACGTC ATACCTATTG GAATTGTGTG GTTCTGTTTG TATTACATCG TTTTCAGATT 11760
 CTTAATTACG AAATTTAATT TCAAAACACC TGGTCGAGAA GATAAAGCTG CAGCACAAACA 11820
 AGTTGAGGCT ACTGAAAGAG CACAACTAT TGTGCTGGT TTGGGAGGCA AAGATAACAT 11880
 35 TGAAATCGTT GACTGTTGTG CAACGAGACT ACGCGTCACA CTTTCATCAA ATGACAAAGT 11940
 CGATAAAGTA TTA¹CTCGAAA GTACTGGTGC CAAAGGTGTA ATCCAGCAAG GCACTGGTGT 12000
 40 GCAAGTAATT TATGGGCCTC ACGTTACAGT TATCAAAAAT GAAATTGAAG AATTGCTCGG 12060
 GGATTAAGAC TAACCGAAAT ATCAACAGAA CTAATGGCAA CGATGTACGA AGTAAGAAGT 12120
 GACATCGTTG CTTTATTTT TAATGTTACA TTTGAAGCAT TAAGTTCATC ATGCACTGTA 12180
 45 GTGAGCCCGC AAATCGCCTC TGCTAGACAA TCATCTTAAT GCTATGATTA AAGCTTAAGT 12240
 GCCAGATTTG AATTTAATTT CAACAACGAC TTTCACTACA TTAAAAATAG GGCCACTCGA 12300
 CACATATAGT TGTATCAAAT AGCCCTTTAT ACAATTTTTT GGGTAAGGTT TTACAATTTT 12360
 50 TGGGATGGTA TAGATTTTAT AAAAAGTTAT TTAAGTTCCT CTGCTTCAGC CATAATATCT 12420
 TTTAATGTTT TAGCTGAATG TCGAACTTG CTTTGTCTT CGTCGTTTAA TGGGATTTCT 12480

55

	TCCTCATATT CGCCTTCTAA TAATGCTGAT ACAGTCAATA CGGCATCTTC ATTTCTGAAA	12600
5	ATCGCTTCAG TAATTCTAGC TAATCCCATT GCAACACCAT AATAAGTGGC ACCTTTAGCT	12660
	TGAATAATGT CATATGCTGC ATCACGTGTT TGAACAAAAA TTTGTTCAAT TTGCGCTTTG	12720
	CCCTCAGGAC GTTGTTC AAG TAATGTCTTC AAAGGTTGAC CCGCAATATT AGCGTGTGAC	12780
10	CATACTGGTA ATTCAGTGTC ACCATGTTCA CCAATAATTT GAGCATCGAC GCTACGTGGC	12840
	GCAACATCGn AcgyTcGCTT AACAAATAATC TAAAGCGTGC AGAGTCTAAA ATTGTACCAG	12900
	AACCTATAAC ACGTTCTTTA GGTAAACCAG AGAATTTCCA TGTTGCATAC GCTAAAATAT	12960
15	CAACAGGATT TGTAGCTACC AAGAAAATAC CATCAAATTT TGATGCCATT ACTTCACCAA	13020
	CAATTGATTT GAATATTTTC AAGTTTTTAG ATACTAAATC TAAACGTGTT TCTCCAGGTT	13080
	TTTGTGCAGC ACCAGCACAG ATGACAATA GATCCGCATC ATGACAATCA CTGTATTCGC	13140
20	CAGCTTTTAC ACGAACTGTT GTTGAGAAAT ATGGTGTGGC ATGTTTTTAAA TCCATAACAT	13200
	CTCCTCGAAC TTTTTCAGTG TCTAAATCAA TGATGACTAA TTCATCAACA ATGCTTTGGT	13260
	TCACTAATGA AAATGCGTAG CTTGAACCTA CTGCACCATT ACCTATTAAT ACAACTTTGT	13320
25	TCCCTTTAAA TTTGTTTATT ACAAAAATC CTTTATGATT AATTCATAA CATACATGTA	13380
	GCTTCAAATA TGTTAGTTTA ATGCTGCTTA TTGACGATAC AAAAGCAAAT AACATCTCT	13440
30	TTTATTTTCA ACGCATAACT TAAAAGGTCA TGTGTCTATC GCTTTTAAAG TTGTGATTTA	13500
	TTTCACATAT AAAATGTAAC ATGCATTAAG TACTGGGTCA ATATTAAATT GTGATTTATT	13560
	TCACATTTTA TTTTAATTTT TACACCTTTT TAATTTGTAT mCGATTACAT CTTAGATGTC	13620
35	TTTAGTCTTC GTACTTCGCC AGTGATTATT TACACTTCA CATTTTATT ATCATGTTTA	13680
	CTTTTTTCTA GGAAAACAAC AATGTTTTTT GAATTAGTCA AATAAATGCG CTCAATCGTC	13740
	GGTGTTGCAA CAGACAATTG TACACAATGC TTATTGATAA GTATTTAAAA AATTAAAAAT	13800
40	GTCATACAAT TATCAAATTT GCCATTTTAT TTATATTTTC TCAAACCAAT TAATTGAATA	13860
	TCGAAATTTT TAGTAGAATA ATCAAAATAT ACAGATTAAA GGAGGAGTAT CATGCTTACA	13920
	GAACAAGAGA AAGACATTAT CAAACAAACG GTGCCTTTAC TTAAAGAGAA AGGGACAGAA	13980
45	ATTACGTCAA TCTTTTATCC AAAAATGTTT AAAGCGCATC CTGAACTTTT AACATGTTT	14040
	AATCAAACGA ACCAAAAACG AGGCATGCAA TCTTCAGCAT TAGCACAAGC TGTAAATGGCC	14100
	GCAGCGGTTA ATATCGATAA CTTAAGTGTT ATTAAACCAG TCATTATGCC AGTCGCATAT	14160
50	AAACACTGCG CACTACAAGT TTATGCTGAA CATTATCCAA TTGTGGGGAA AAATTTATTA	14220
	AAAGCCATTC AAGACGTGAC AGGATTAGAA GAAAATGACC CTGTCATTCA AGCTTGGGCA	14280

55

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8779 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```

GGTATTTTnG GAnGGGTACC TAAAGCAATT CCGGCAAAGG GTnAATCCAG GTACCGAAAT      60
GGACTTCCCG TTATCGATAA TACCGACATA TATTGTGACA AGTAGATTTT ATGGACATTT      120
AGGCTTACTT TTACTTGTGA TAATTGCATG TATGTTTACT GGTATTTAtC CaTCaATACA      180
TATCATTCaA TTATTGATAT ATGTACCGTT TTGTTTTTTC TTAActGCct CGGTGACGTT      240
ATTAACATCA ACACTCGGTG TGTTAGTTAG AGATACACAA ATGTTAATGC AAGCAATATT      300
AAGAATATTA TTTTACTTTT CACCAATTTT GTGGCTACCA AAGAACCATG GTATCAGTGG      360
TTTAATTCAT GAAATGATGA AATATAATCC AGTTTACTTT ATTGCTGAAT CATACCGTGC      420
AGCAATTTTA TATCACGAAT GGTATTTTCAT GGATCATTGG AAATTAATGT TATACAATTT      480
CGGTATTGTT GCCATTTTCT TTGCAATTGG TCGGTACTTA CACATGAAAT ATAGAGATCA      540
ATTTGCAGAC TTCTTGTAAT ATATTTATAT GACGAAACCC CGCTAACCAT TAATAAATGG      600
AAGTGGGGTT CATTTTTGTT TATAATTTAA GTAAATAACA TATTAAGTTG GTGTATTATG      660
AACGTTTTTA TAAAGAAATT TTATCATTTG GTAGTTCGAA TACTTTCTAA AATGATTACG      720
CCTCAAGTGA TTGATAAACC GCATATCGTA TTTATGATGA CTTTTCAGAG AGATATTAAG      780
CCTATCATCA AAGCATTAAA TAATTCGTCTG TATCAGAAAA CTGTTTAAAC AACACCAAAA      840
CAAGCGCCTT ATTTATCTGA ACTTAGCGAC GATGTTGATG TGATAGAAAT GACTAATCGA      900
ACATTGGTAA AACAAATTAA GGCTTTGAAA AGCGCGCAGA TGATTATTAT CGATAATTAT      960
TACCTATTGC TAGGTGGATA TAATAAGACT TCTAATCAAC ACATTGTTCA AACGTGGCAT     1020
GCAAGTGGTG CATTAAAAAA CTTTGGCTTA ACAGATCATC AAGTCGATGT GTCTGACAAG     1080
GCAATGGTTC AGCAGTACCG TAAAGTTTAT CAAGCGACGG ATTTTACTTT AGTGGGTTGT     1140
GAACAAATGT CACAATGTTT TAAACAGTCT TTAGGTGCAA CAGAAGAGCA AATGCTGTAT     1200
TTTGGGCTTC CGAGAATTAA TAAATATTAC ACAGCTGATA GAGCAACGGT TAAGGCAGAG     1260
TTAAAGGATA AATATGGAAT TACAAATAAG TTGGTATTAT ATGTACCAAC ATATAGAGAA     1320
GATAAAGCAG ATAATAGGGC TATTGATAAA GCTTATTTTG AAAAATGTTT ACCAGGATAT     1380

```

	ATCGACACGT CTACATTAAT GCTAATGTCA GATATAATTA TTAGCGACTA TAGTTCGCTG	1500
	CCAATAGAAG CTAGCTTGTT AGATATTCCA ACTATATTTT ATGTGTATGA TGAAGGAACA	1560
5	TATGATCAGG TGAGAGGCCT GAATCAATTT TACAAAGCAA TACCGGATAG CTACAAAGTG	1620
	TATACTGAAG AAGATTTAAT AATGACGATA CAAGAAAAAG AACATCTATT AAGTCCGTTA	1680
	TTTAAAGATT GGCATAAGTA TAATACTGAT AAAAGTTTAC ATCAGCTCAC AGAATATATA	1740
10	GATAAGATGG TGACAAAATG AGGTTTACGA TAATCATACC TACATGTAAT AATGAGGCAA	1800
	CAATTCGACA ATTGTTAATA TCTATTGAGA GTAAAGAACA CTATAGAATC CTTTGTATTG	1860
	ATGGTGGTTC TACTGATCAA ACAATTCCTA TGATTGAACG GTTACAAAGA GAACTCAAGC	1920
15	ATATTTCAAT AATACAATTA CAAAATGCTT CGATAGCTAC GTGTATTAAT AAAGGTTTGA	1980
	TGGATATCAA AATGACAGAT CCACATGATA GTGACGCATT TATGGTCATA AAACCAACAT	2040
20	CAATCGTATT GCCAGGTAAA TTAGATAGGT TAACTGCTGC TTTCAAAAAT AATGATAATA	2100
	TTGATATGGT AATAGGGCAG CGAGCTTACA ATTACCATGG TGAATGGAAA TTGAAAAGTG	2160
	CTGATGAGTT TATTAAAGAC AATCGAATCG TTACATTAAC GGAACAACCA GATTGTATTAT	2220
25	CAATGATGTC TTTTGACGGA AAGTTATTCA GTGCTAAATT TGCTGAATTA CAGTGTGACG	2280
	AAACTTTAGC TAACaCATAC AATCACGCAA TACTTGTCaA GGCGATGCAA AAAGCTACGG	2340
	ATATACATTT AGTTTCACAG ATGATTGTCTG GAGATAACGA TATAGATACA CATGCTACAA	2400
30	GTAACGATGA AGATTTTAAT AGATATATCA CAGAAATTAT GAAAATAAGA CAACGAGTCA	2460
	TGGAAATGTT ACTATTACCT GAACAAAGGC TATTATATAG TGATATGGTT GATCGTATTT	2520
	TATTCAATAA TTCATTAAAA TATTATATGA ACGAACACCC AGCAGTAACG CACACGACAA	2580
35	TTCAACTCGT AAAAGACTAT ATTATGTCTA TGCAGCATTG TGATTATGTA TCGCAAAACA	2640
	TGTTTGACAT TATAAATACA GTTGAATTTA TTGGTGAGAA TTGGGATAGA GAAATATACG	2700
40	AATTGTGGCG ACAAACATTA ATTCAAGTGG GCATTAATAG GCCGACTTAT AAAAAATTCT	2760
	TGATACAACT TAAAGGGAGA AAGTTTGAC ATCGAACAAA ATCAATGTGA AAACGATAAC	2820
	GTGTACATTG ATGACCATAA ACTGCAATCC TATGATGTGA CAATATGAGG AGGATAACTT	2880
45	AATGAAACGT GTAATAACAT ATGGCACATA TGACTTACTT CACTATGGTC ATATCGAATT	2940
	GCTTCGTCGT GCAAGAGAGA TGGGCGATTA TTTAATAGTA GCATTATCAA CAGATGAATT	3000
	TAATCAAATT AAACATAAAA AATCTTATTA TGATTATGAA CAACGAAAAA TGATGCTTGa	3060
50	ATCAATACGC TATGTCGATT TAGTCATTCC AGAAAAGGGC TGGGGACAAA AAGAAGACGA	3120
	TGTCGAAAAA TTTGATGTAG ATGTTTTTGT TATGGGACAT GACTGGGAAG GTGAATTCGA	3180

55

	TAAAATCAAA CAAGAATTAT ATGGTAAAGA TGCTAAATAA ATTATATAGA ACTATCGATA	3300
	CTAAACGATA AATTAACTTA GGTATTATA AAATAAATAT AAAACGGACA AGTTTCGCAG	3360
5	CTTTATAATG TGCAACTTGT CCGTTTTTAG TATGTTTTAT TTTCTTTTTC TAAATAAACG	3420
	ATTGATTATC ATATGAACAA TAAGTGCTAA TCCAGCGACA AGGCATGTAC CACCAATGAT	3480
	AGTGAATAAT GGATGTTCTT CCCACATACT TTTAGCAACA GTATTTGCCT TTTGAATAAT	3540
10	TGGCTGATGA ACTTCTACAG TTGGAGGTCC ATAATCTTTA TTAATAAATT CTCTTGATA	3600
	GTCCGCGTGT ACTTTACCAT CTCGACTAC AAGTTTATAA TCTTTTTTAC TAAAATCACT	3660
	TGGTAAAACA TCGTAAAGAT CATTTTCAAC ATAATATTTT TTACCATTTA TCCTTTGCTC	3720
15	ACCTTTAGAC AATATTTTTA CATATTTATA CTGATCAAAT GAGCGTTCCA TTAATGCATT	3780
	CCCCATCATA TTACGTTGCT TCTCGCCACC AAGGTTTTTA TAGTCTCCTG CACCCATGAT	3840
20	AACCTGATT AATTCTAAATT TACCTCGTTT GGTAGTAATC GTATGGTTGT AATTGCTGT	3900
	ATCACTTGAT CCAGTTTTTA AACCATCTGT ACCCGGCAAA CTCATTTTTG CACCTTCCAA	3960
	TGAAAAGTTG AATGTGTAAT ACGTAACTGC ATGCGTTGTT GGTGCTAACT GCTTTGTAAA	4020
25	GTCTAATATT TTAGGTGTCT CTTTAATCAC GTGTAAATCT AAAATGGCAT AGTCTCTAGC	4080
	AGTCGTTACA GTACGTTCTT GGTCTTTATA CTTTGTGGT GCAAATGTAC GTAATCTTGA	4140
	ATTTTCAGCA CCCGTTGGAT TGACGAAATG TGTATTTTTC ATTCCGATAG CTTTAGCTTT	4200
30	GTTATTCATT AAATCAACGA AATCGCTGGT GTTTTTTGAA ACCTTCTTAG CTAAAATTAA	4260
	TGCCGCGGCA TTAGTAGAAT TAGATACTGT AATTTGTAAT AGGTCTGCGA TTGTCCATAC	4320
	TTGTCCAGGA TATAGTTTCG TATTACTCAA CTCAGGTAGT GTAGACATAA TATATTCTTT	4380
35	GTTGCTCATT GTGACTGTGT CATCAAGTGA AAGCTGCCCC TTATTTACAG CTTCCAATGT	4440
	TAAGTACATT GTCATTAATT TAGTCATAGA CGCTGGATTC CACTTAGTAT CGATATTGTA	4500
	TTGATACAGT AATTGTCCAG TTTGACTTAC ATTAACAGCA CTCGTCGGTT CGTATGCAGC	4560
40	CGACAAACCT GCATAACCAT ATTGATTGTC TGCTTGATCA GGGGTTACGT CACTGTTAGT	4620
	AGCTTGTGCA TATGGTGTCA TAATACTTAA TGTAAACAT AAAATGATGA TAATAGATAT	4680
45	TAAATTTTTT ATAAAGCGTT AATCTTCCCT TTTCCAATTC TTAAATATTC CCTAAAAGCA	4740
	ATGGTTATTC CTAATTACGG AAATCATTGC TAATCACTT CACCTTAATT AAATTGTTGA	4800
	AAATAAGTT TTCTGCAGTT AATTTGAAAA ATAATGCAAA TATATTACGT GTGTAGCTAA	4860
50	AGGTGTTATA ATGTTTGTAC GAAGAGCAAA CTTACTCAAA AGCGATTAAAT TTTCATGTTT	4920
	TAATATAAAG ACTTTGAGAA GTTATTACAA AAAATGCAAT AGAAATATTC TATCATATAA	4980

55

	AAGTATATGA TAGAAATGCA TGTATCTATC TAAATGAATT AACTATAAAT TTCAAACAGA	5100
	AGAGGTAAAA CTATGAAACG AGAAAATCCA TTGTTTTTCT TATTTAAAAA ACTATCATGG	5160
5	CCAGTGGGTC TTATCGTTGC AGCTATCACT ATTTTCATCAC TAGGGAGCTT AAGTGGAATA	5220
	TTAGTGCCAC TGTTTACTGG ACGAATTGTA GATAAAATTTT CCgTGAGCCA TATCAATTGG	5280
	AATCtAATCG CATTATTTGG TGGTATCTTT GTCATCAATG CTTTATTAAG CGGATTAGGT	5340
10	TTATATTTAT TAAGTAAAT TGGTGAAAAG ATTATTTATG CGATACGCTC AGTTTTATGG	5400
	GAGCATATCA TACAATTAAA AATGCCAATC TTTGACAAAA ATGAAAGTGG TCAATTAATG	5460
	AGTCGATTAA CTGACGATAC GAAAGTGATA AATGAATTTA TTTCACAAAA GCTACCTmAC	5520
15	TTATTACCAT CAATCGTTAC ATtAGTTGGG TCACTAATCA TGTTATTTAT TTTAGATTGG	5580
	AAAATGACAT TATTAACATT TATAACGATA CCGATATTCG TTTTaATTAT GATTCTCTTA	5640
20	GGTCGTATTA TGCAAAAGAT ATCGACAAGT ACACAATCTG AAATTGCAAA CTTCAGTGGT	5700
	TTGTTAGGGC GTGTCCTAAC TGAAATGCGT CTTGTTAAAA TATCAAATAC AGAGCGTCTT	5760
	GAATTAGATA ATGCACATAA AAATTTGAAT GAAATATATA AATTAGGTTT AAAACAGGCT	5820
25	AAAATGCGG CAGTTGTACA ACCAATTTCA GGTATAGTTA TGTGCTAAC AATTGCAATT	5880
	ATTTTAGGTT TTGGTGCATT AGAAATTGCG ACTGGTGCAA TCACTGCAGG TACATTAATT	5940
	GCAATGATAT TTTATGTTAT TCAGTTATCT ATGCCTTTAA TCAATCTTTC CACGTTAGTT	6000
30	ACAGATTATA AAAAGGCAGT CCGTGCAAGT AGTAGAATAT ACGAAATCAT GCAAGAACCT	6060
	ATTGAACCGA CAGAAGCTCT TGAAGATTCT GAAAATGTAT TAATTGATGA CGGTGTATTG	6120
	TCATTTGAAC ATGTAGACTT TAAATATGAT GTGAAGAAAA TATTAGATGA TGTGTCGTTT	6180
35	CAAATCCAC AAGGTCAAGT GAGTGCTTTT GTAGGCCCTT CTGGGTCTGG TAAAAGTACG	6240
	ATATTTAATC TGATAGAACG TATGTATGAA ATTGAGTCAG GTGATATTAA ATATGGCCTT	6300
	GAAAGTGTCT ATGATATCCC GTTATCTAAG TGGCGACGCA AAATTGGATA TGTTATGCAA	6360
40	TCAAATTCGA TGATGAGTGG TACAATTAGA GACAATATTT TATACGGAAT TAATCGTCAT	6420
	GTTTCAGATG AAGAACTTAT TAATTATGCT AAATTAGCGA ACTGTCATGA TTTTATCATG	6480
45	CAATTTGATG AAGGATATGA CACGCTTGTA GGTGAACGAG GATTGAACT GTCTGGCGGA	6540
	CAACGTCAAC GTATTGATAT TGCTAGAAGT TTTGTTAAAA ATCCTGATAT TTTGTTACTT	6600
	GATGAAGCAA CAGCTAATCT CGATAGTGAA AGTGAATTGA AAATTCAAGA AGCTTTAGAA	6660
50	ACATTGATGG AAGGTAGAAC AACGATTGTC ATTGCGCATC GTTTGTCTAC AATTAAAAAA	6720
	GCCGGTCAAA TTATATTCTT AGACAAAGGA CAGGTAACAG GTAAAGGTAC GCATTCAGAA	6780

55

	TTTTATATAT ATAAGTAAGC TTGGAGCAAA TACACATATA CCATCGAGGA AATTAAAGTG	6900
	TGGCACATTG ATGGATATAG ATGTTAATAA ATTGCTTCAA GCTTTTGTCT ATTTTAAATC	6960
5	ATTTGAGAAG TTACGACATA ATAATTCTTA AATTAATGAA ATCGATATTT TAAGAAAAAA	7020
	ATGCTCATGG TATAATACAA GTTATAAGCA AACATACATA TATTAAATAC TGTAGCCACG	7080
	AGTCATAATT CTTTCATATTT TACATAGCAA TTAACTGAT TTTAGAGTCC ACGGTACAGA	7140
10	AGTTTGATAT TTCAATGTTT CTAAATTTTT AAAAAATTAA ATCATAGGTG GGTGCCAAAT	7200
	GTTTTTATTA ATCAACATTA TTGGTCTAAT TGTATTTCTT GGTATTGCGG TATTATTTTC	7260
	AAGAGATCGC AAAAATATCC AATGGCAATC AATTGGGATC TTAGTTGTTT TAAACCTGTT	7320
15	TTTAGCATGG TTCTTTATTT ATTTTGATTG GGGTCAAAAA GCAGTAAGAG GAGCAGCCAA	7380
	TGGTATCGCT TGGGTAGTTC AGTCAGCGCA TGCTGGTACA GGTTTTGCAT TTGCAAGTTT	7440
20	GACAAATGTT AAAATGATGG ATATGGCTGT TGCAGCCTTA TTCCAATAT TATTAATAGT	7500
	GCCATTATTT GATATCTTAA TGTACTTTAA TATTTTACCG AAAATTATTG GAGGTATTGG	7560
	TTGGTTACTA GCTAAAGTAA CAAGACAACC TAAATTCGAG TCATTCTTTG GGATAGAAAT	7620
25	GATGTTCTTA GGAAATACTG AAGCATTAGC CGTATCAAGT GAGCAACTAA AACGTATGAA	7680
	TGAAATGCGT GTATTAACAA TCGCAATGAT GTCAATGAGC TCTGTATCGG GAGCTATTGT	7740
	AGGTGCGTAT GTACAAATGG TACCAGGAGA ACTGGTACTA ACGGCAATTC CACTAAATAT	7800
30	CGTTAACGCG ATTATTGTGT CATGCTTGTT GAATCCAGTA AGTGTTGAAG AGAAAGAAGA	7860
	TATTATTTAC AGTCTTAAAA ACAATGAAGT TGAACGTCAA CCATTCTTCT CATTCTTG	7920
	AGATTCTGTA TTAGCAGCAG GTAAATTAGT ATTAATCATC ATCGCATTG TTATTAGTTT	7980
35	TGTAGCGTTA GCTGATCTAT TTGATCGTTT TATCAATTTG ATTACAGGAT TGATAGCAGG	8040
	ATGGATAGGC ATAAAAGGTA GTTTCGGTTT AAACCAAATT TTAGGTGTGT TTATGTATCC	8100
40	ATTTGCGCTA TTAICTGGTT TACCTTATGA TGAAGCGTGG TTGGTAGCAC AACAAATGGC	8160
	TAAGAAAATT GTTACAAATG AATTTGTTGT TATGGGTGAA ATTTCTAAAG ATATTGCATC	8220
	TTATACACCA CACCATCGTG CGGTTATTAC AACATTCTTA ATTTCAATTG CAAACTTCTC	8280
45	AACGATTGGT ATGATTATCG GTACATTGAA AGGCATTGTT GATAAAAAGA CATCAGACTT	8340
	TGTATCTAAA TATGTACCTA TGATGCTATT ATCAGGTATC CTAGTTTCAT TATTAACAGC	8400
	AGCTTTCGTT GGTTTATTTG CATGGTAATA TGTCGAAGAG TGAATATGAT AATACATTTT	8460
50	AACTAATAAA TATGTCCAGG CATGTCGTCT ATTGATATAG GTGAGATGCT TGGACTTTTT	8520
	TATTATTGAT ATAAAGGTAT nTAAATATTT TTAAAGTTAC CGAAATTGAA GCATTATAAA	8580

55

GACAGTAAGG ACTAGGTACA GTCATAGTAC TTCGAGCAAA ATTTGTTTGT TTATTATAAA 8700
 CAACACAAAG GAGATAACTT CTCTAATGAA GAAGTTAAAA ACATTATAGC AGACAATGAA 8760
 5 ATGAAAGTAA ATTAAAAAT 8779

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31096 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GTTGCAGTAG TCAAAGAATT AAACAAGGTG AAGGCGTGTA GCTTGACAC CCGAAAATGT 60
 20 GCGTAAGTTA aCGGATGCAG GACATAAAGT AATTGTTGAA AAAAATGCTG GCATTGGTTC 120
 AGGATTTTCT AACGATATGT ATGAAAAAGA AGGCGCTAAG ATCGTAACTC ACGAACAAGC 180
 ATGGGAAGCT GATCTTGTTA TCAAAGTAA AGAACCTCAT GAAAGCGAAT ATCAATATTT 240
 25 CAAAAAGAAT CAAATTATCT GGGGATTTT ACATCTAGCA TCTTCAAAG AAATAGTAGA 300
 AAAAATGCAA GAAGTTGGTG TAACTGCGAT TAGTGGTGAA ACCATTATAA AAAATGGAAA 360
 AGCAGAATTA TTAGCGCCAA TGAGTGCTAT AGCAGGTCAA CGCTCAGCAA TTATGGGAGC 420
 30 TTACTACTCT GAAGCACAA ACATGGTCA AGGTACTTTA GTGACTGGTG TACATGAAAA 480
 TGTGGATATA CCTGGTAGTA CATATGTGAT TTTCGGTGGT GGAGTAGCAG CAACAAATGC 540
 AGCAAATGTT GCCTTGGGAC TAAATGCTAA AGTAATCATT ATCGAGTTAA ACGATGACCG 600
 35 CATTAAATAT CTTGAAGATA TGTATGCAGA AAAAGATGTC ACAGTAGTCA AATCAACACC 660
 AGAAATTTTA GCAGAACAAA TTAAGAAAGC AGATGTATTT ATTTCTACAA TTTTAATTTT 720
 40 AGGTGCGAAA CCGCCAAAAT TGGTTACTCG TGAGATGGTT AAATCAATGA AAAAAGGTTC 780
 AGTATTAATC GATATAGCTA TTGACCAAGG TGGAACCTATT GAAACAATTA GACCAACTAC 840
 AATTTCTGAT CCAGTGTATG AAGAAGAAGG TGTGATTCAT TATGGTGATC CAAATCAACC 900
 45 AGGAGCAGTC CCAAGAACTT CAACAATGGC ATTAGCACAA GGAAATATTG ATTATATATT 960
 AGAAATTTGT GACAAAGGCT TAGAACAAGC AATTAAAGAT AATGAAGCCT TAAGTACTGG 1020
 TGTAACATT TACCAAGGAC AAGTGACAAA TCAAGGATTA GCTTCATCAC ATGACCTAGA 1080
 50 TTATAAGAA ATATTAAATG TTATCGAATA GATAGTAATT TAAATGAAAT TGAGTGAAAT 1140
 GAATATTTTA AATATAGCAT TATAGTTTGG ACTAAAAATT TACAAAACGG AAGGATGTAA 1200

	TCGAAGAAGC TAAAGCAAGC ATTAAACCAT TTATTCGTCG AACACCTCTA ATTAAATCAA	1320
	TGTATTTAAG CCAAAGTATA ACTAAAGGGA ATGTATTTCT AAAATTAGAA AATATGCAAT	1380
5	TCACAGGATC TTTTAAATTT AGAGGCGCTA gCAATnAAAA TTAATCACTT AACAGATGAA	1440
	CAAAAAGAAA AAGGCATTAT CGCAGCATCT GCTGGGgAAC CATGCACAAG GTGTTGCTTT	1500
	AACAGCTAAA TTATTAGGCA TTGATGCAAC GATTGTAATG CCTGAAACAG CACCACAAGC	1560
10	GAAACAACAA GCAACAAAAG GCTATGGGGC AAAGGTTATT TTAAGGTA AAAACTTTAA	1620
	CGAACTAGA CTTTATATGG AAGAATTAGC GAAAGAAAAT GGCATGACAA TCGTTCATCC	1680
	ATATGACGAT AAGTTTGTAA TGGCAGGCCA AGGAACAATT GGTTTAGAAA TTTTAGATGA	1740
15	TATTTGGAAT GTGAATACAG TCATCGTACC AGTTGGCGGT GGAGGATTAA TTGCAGGTAT	1800
	TGCCACCGCA TTAATCAT TTAACCTTC AATTCATATT ATCGGTGTTT AATCTGAGAA	1860
20	TGTTTCATGGT ATGGCTGAGT CTTTCTATAA GAGAGATTTA ACTGAACATC GAGTGGATAG	1920
	CACAAATAGCA GATGGTTGTG ATGTAAAAGT TCCTGGTGAA CAAACATATG AAGTAGTTAA	1980
	ACATTTAGTA GATGAATTTA TTCTTGTTAC TGAAGAAGAA ATTGAACATG CTATGAAAGA	2040
25	TTTAATGCAG CGTGCCAAAA TTATTACTGA AGGTGCAGGC GCATTACCAA CAGCTGCAAT	2100
	TTTAAGTGGA AAAATAAACA ATAAATGGCT TGAAGATAAA AATGTTGTTG CATTAGTTTC	2160
	AGGCGGGAAT GTTGACTTAA CTAGAGTTTC AGGTGTCATT GAACATGGAC TGAATATTGC	2220
30	AGATACAAGC AAGGGTGTGG TAGGTAAAAA CATTTAATCT TAAAAATGAG GTGTAATTAT	2280
	GTCAAATGGT AAAGAATTAC AAAAAAATAT AGGTTTCTTC TCAGCGTTTG CTATTGTTAT	2340
	GGGGACAGTT ATTGGTTCAG GAGTATTCTT TAAATATCA AACGTAACAG AAGTAACAGG	2400
35	AACAGCAGGA ATGGCCTTGT TTGTATGGTT CCTAGGCGGC ATCATTACCA TTTGTGCGGG	2460
	GTTAAACAGCA GCAGAACTTG CTGCTGCAAT CCCTGAAACA GGTGGCTTAA CGAAGTATAT	2520
	AGAATATACA TACGGTGATT TCTGGGGCTT CCTATCAGGT TGGGCGCAAT CATTTATTTA	2580
40	TTTTCCAGCT AACGTAGCAG CATTGTCTAT CGTATTTGCG ACACAGCTAA TTAATTTATT	2640
	CCATTTATCT ATAGGTTTCGT TAATACCAAT AGCAATCGCA TCTGCGTTAT CTATTGTGTT	2700
45	GATAAATTTT CTAGGTTCAA AAGCAGGCGG AATTTTACAA TCAGTTACTT TAGTAATTAA	2760
	ACTGATTCCA ATCATCGTTA TTGTAATTTT TGGTATTTTT CAATCTGGAG ATATCACTTT	2820
	TTCAATTAAT CCAACTACAG GTAATTCaGG AAATGGCTTC TTACAGCAA TTGGTAGTGG	2880
50	TTTATTAGCA ACTATGTTTG CATATGATGG TTGGATTCAT GTAGGAAATG TTGCGGGGGA	2940
	ACTTAAAAAT CCTAAACGCG ATTTACCTTT AGCGATTTCA GTTGGTATCG GTTGTATTAT	3000

55

	TGGTAATTTA AATGCAGCTT CAGATACATC AAAAATATTA TTTGGTGAAA ATGGCGGTAA	3120
	GATTATTACA ATCGGTATAT TAATTTCTGT TTATGGTACG ATCAATGGCT ATACTATGAC	3180
5	TGGTATGCGC GTACCATATG CAATGGCTGA AAGAAAATTA TTGCCATTTA GCCATTTATT	3240
	CGCAAAATTA ACAAATCTG GCGCACCATG GTTTGGCGCA ATTATACAAC TTATAATCGC	3300
	TATCATCATG ATGTCAATGG GAGCATTGTA TACAATTACA AATATGTTAA TCTTTGTTAT	3360
10	TTGGTTGTTT TATTGTATGT CATTTGTTGC GGTAATAATT TTAAGAAAAC GTGAACCAA	3420
	TATGGAACGA CCATATAAAG TACCGTTATA TCCGATCATA CCTTTAATTG CTATTTTGGC	3480
	AGGATCATTT GTATTAATTA ATACACTGTT TACACAATTT ATATTAGCAA TCATTGGAAT	3540
15	TCTAATAACA GCACTTGGA TACCAGTTTA TTACTATAAA AAGAAACAAA AAGCAGCATA	3600
	AGGTAAGATA ACTAGCATTG AGAATAAATG GATGGACTAC TAATAAATTT AAAGTTTTAC	3660
20	ACATTAATAA CAAAAACCAT TCAATTATTC TATGGAACAG ACAAATTTCT GTTATGGAAT	3720
	TTGTCTGTTT TTCAAAAGTA TAGGGAGGCA AATAGAGATG GAAAAGCCGT CAAGAGAGGC	3780
	ATTTGAAGGC AATAATAAGT TGTTAATAGG AATTGTTCTA AGTGAATAA CGTTTTGGCT	3840
25	ATTTGCACAA TCATTGGTTA ATGTTGTACC AATACTTGAA GATAGTTTCA ATACAGATAT	3900
	TGGAACGGTT AATATCGCCG TTAGTATAAC TGCTTTATTT TCAGGAATGT TTGTAGTAGG	3960
	AGCAGGTGGT CTTGCTGATA AATATGGCAG AATTAACTC ACGAACATTG GTATTATCTT	4020
30	AAATATATTA GGTTCAATTAT TAATCATTAT TTCAAATATT CCTTTATTAC TTATTATAGG	4080
	AAGATTAATT CAAGGACTTT CAGCAGCATG TATTATGCCT GCAACTTTGT CTATTATTAA	4140
	GTCATATTAC ATTGGGAAAG ATAGACAACG CGCTTTAAGT TATTGGTCAA TTGGCTCATG	4200
35	GGGCGGCTCT GGTGTTTGTT CATTTTTTGG AGGTGCAGTT GCAACGCTTT TAGGTTGGCG	4260
	TTGGATTTTC ATCCTATCAA TTATAATTTT ATTAATTGCA CTGTTTCTTA TTAAAGGCAC	4320
	ACCTGAAACT AAATCTAAAT CGATTTCTCT AAATAAATTT GACATTAAAG GTCTGGTTCT	4380
40	TTTAGTCATT ATGTCCTCA GTTTAAATAT TTAAATTACT AAAGGATCAG AATTAGGTGT	4440
	AACCTCACTT CTTTTATTA CTTTATTAGC TATTGCAATT GGATCTTTTA GTTTATTAT	4500
45	AGTTCTTGAA AAGCGTGCTA CAAATCCTTT AATCGATTTT AAATTATTTA AAAATAAAGC	4560
	TTACACAGGT GCAACAGCTT CAACTTTTT GTTAAATGGT GTTGCAAGGAA CATTAAATAGT	4620
	AGCCAACACA TTTGTTCAAA GAGGTTTAGG ATATTCTTCA TTGCAAGCAG GAAGTTTATC	4680
50	AATCACTTAT TTAGTAATGG TACTAATTAT GATTCGTGTT GGTGAAAAGT TACTTCAAAC	4740
	ACTCGGATGC AAGAAACCAA TGTTAATTGG AACAGGAGTT CTTATTGTCG GAGAATGTCT	4800

55

	ATTCTTTGGT TTAGGACTAG GGATATATGC TACACCATCA ACAGATACAG CAATTGCAAA	4920
	TGCACCGTTA GAAAAAGTAG GCGTTGCTGC AGGTATCTAT AAAATGGCTT CTGCATTAGG	4980
5	TGGAGCATTT GGCGTCGCAT TGAGTGGTGC AGTATATGCA ATCGTATCAA ATATGaCAAA	5040
	CATTTATACA GGTGcAAATGa TTGnCATTAT GGTTaAATGC AGGTATGGGa ATATTATCaT	5100
	TCGTTATCAT TTTGtTACTT GTGcCTAAAC mAAACGACAC TCAATTATGA TAATTGAGAA	5160
10	TTAAATTGAA ATCATACAAG TCGCTACAAT ATTAAACAAA AATATAAACC GATTCTTATG	5220
	TGTCATTATT TTAAATGAAC ATAGGGATTG GTTTTTTATT ACTCTTTTAC GCTACTTTAT	5280
	TTATAATTAT TATAAATTGT CACAAATTCA ATTTACCTTA CAATATATTT TGTGTTATTA	5340
15	TATTCTGGAG CATAAATAAA TTGTTCAACA CATAGTTGTA ATGTGTTTCA ATACTTTTTG	5400
	GATAGATTGC GAAATTGTAT TGAATCGTCA TCGTTTTAAA TTTTAAATG AGAATGGAAT	5460
20	GAGCATTACA ATACACAAGC AATCAAAAGT AAATACATTC ACAACACAAC AGAGACATAA	5520
	CAACAAGATA AGGAGTGAAC AATAGCTGTG AATTATCGTG ATAAAAATCA AAAGTTTAGT	5580
	ATTCGTAAAT ATACAGTTGG TACATTTTCA ACTGTCATTG CGACATTGGT ATTTTtagGA	5640
25	TTCAATACAT CACAAGCACA TGCTGCTGAA ACAAATCAAC CAGCAAGCGT GGTTAAACAG	5700
	AAACAACAAA GTAATAATGA ACAGACTGAG AATCGAGAAT CTCAAGTACA AAATTCTCAA	5760
	AATTCACAAA ATGGTCAATC ATTATCTGCT ACTCATGAAA ATGAGCAACC AAATATTAGT	5820
30	CAAGCTAATT TAGTAGATCA AAAAGTAGCG CAATCATCTA CTAATAATGA TGAACAACCA	5880
	GCATCTCAAA ATGTAAATAC AAAGAAAGAT TCGGCAACGG CTGCGACAAC ACAACCAGAT	5940
	AAAGAACAAA GTAAGCATAA ACAAACGAA AGTCAATCTG CTAATAAAAA TGGAAACGAC	6000
35	AATAGAGCGG CTCATGTAGA AAATCATGAA GCAAATGTAG TAACAGCTTC AGATTCATCT	6060
	GATAATGGTA ACGTACAACA TGACCGAAAT GAATTACAAG CGTTTTTTGA TGCAAATTAT	6120
	CATGATTATC GCTTTATTGA CCGTGAAAAT GCAGATTCTG GCACATTTAA CTATGTAAAA	6180
40	GGCATTTTTG ATAAGATTAA TACGTTATTA GGCAGTAATG ATCCAATAAA CAATAAAGAC	6240
	TTGCAACTTG CATACAAAGA ATTGGAACAA GCTGTTGCTT TAATTCGTAC AATGCCTCAA	6300
45	CGTCAACAGA CTAGCCGACG TTCAAATAGA ATTCAAACGC GTTCGGTTGA GTCAAGAGCT	6360
	GCAGAGCCTA GATCAGTATC AGACTATCAA AATGCAAATT CATCATATTA TGTGAAAAT	6420
	GCTAATGATG GTTCGGGCTA TCCTGTTGGT ACATATATCa ATGCTTCTAG TAAAGGGGCG	6480
50	CCATATAATT TACCAACTAC ACCATGGAAT ACATTGAAGG CCTCTGACTC AAAGGAAATT	6540
	GCTCTTATGA CAGCGAAACA AACTGGAGAC GGGTACCAAT GGGTTATTAA GTTTAATAAA	6600

55

EP 0 786 519 A2

	GTAGGAAGAA CTGACTTTGT AACAGTTAAT TCAGATGGAA CAAATGTACA ATGGAGTCAT	6720
	GGAGCAGGAG CAGGTGCAAA TAAACCACTT CAACAAATGT GGGAATATGG AGTAAATGAT	6780
5	CCTCATCGTT CACATGACTT TAAAATAAGA AATAGAAGTG GCCAAGTAAT ATATGACTGG	6840
	CCAACTGTCC ATATTTATTC TTTAGAAGAT TTATCTAGAG CGAGTGATTA TTTTAGTGAA	6900
	GCTGGAGCGA CACCTGCTAC TAAAGCTTTT GGTAGACAAA ATTTTGAATA TATTAATGGT	6960
10	CAAAAACCTG CTGAATCACC GGGTGTTCCT AAAGTTTATA CTTTCATCGG TCAAGGTGAT	7020
	GCAAGTTATA CAATTTTATT TAAACACAA GGTCCAACCTG TTAATAAATT GTACTATGCA	7080
	GCAGGTGGGC GTGCTTTAGA GTACAATCAA TTATTTATGT ACAGTCAACT ATACGTCGAA	7140
15	TCAACGCAAG ACCATCAACA ACGTCTTAAT GGTTTAAGAC AAGTGGTTAA TCGTACATAT	7200
	CGCATAGGTA CAACTAAACG TGTAGAAGTG AGTCAAGGAA ATGTACAAAC GAAAAAGGTA	7260
20	TTAGAAAGTA CAAACCTAAA TATAGATGAT TTTGTTGATG ATCCTTTAAG TTATGTTAAG	7320
	ACGCCGAGTA ATAAAGTGTT AGGATTTTAT TCGAATAATG CAAATACTAA TGCTTTTAGA	7380
	CCGGGTGGAG CCCAACAATT AAATGAATAT CAATTAAGTC AATTATTTAC TGATCAAAAA	7440
25	TTACAAGAAG CAGCAAGAAC TAGAAACCCA ATAAGATTAA TGATTGGTTT CGACTATCCT	7500
	GATGCTTATG GTAATAGTGA AcTTTAGTTC CTGTTAACTT AACGGTATTA CCTGAAATCC	7560
	AACATAATAC TaAATTCTTT AAAAATGACG ATACTCAAAA TATTGCTGAA AAACCATTTT	7620
30	CAAAACAAGC TGGGCATCCA GTTTCTATG TATATGCAGG TAACCAAGGG AATGCTTCCG	7680
	TGAATTTAGG TGGTAGCGTA ACATCTATTC AACCATTACG TATTAATTTA ACAAGTAATG	7740
	AGAATTTTAC AGATAAAGAT TGGCAAATTA CAGGTATTCC GCGTACATTA CACATTGAAA	7800
35	ACTCGACAAA TAGACCTAAT AATGCCAGAG AACGCAATAT TGAACCTGTT GGTAACCTAT	7860
	TACCGGGGGA TTACTTTGGA ACGATACGTT TTGGACGTAA AGAACAATTA TTCGAAATTC	7920
	GTGTTAAACC ACATACACCA ACAATTACAA CGACAGCTGA GCAATTAAGA GGTACAGCAT	7980
40	TACAAAAAGT GCCTGTTAAT ATTTCTGGGA TACCGTTGGA TCCATCGGCA TTGGTTTATT	8040
	TAGTTGCACC AACAAATCAA ACTACGAATG GTGGTAGTGA GGCAGATCAA ATACCATCTG	8100
45	GTTATACGAT ACTTGCGACT GGTACACCTG ATGGGGTGCA TAATACAATT ACTATACGAC	8160
	CGCAAGATTA TGTTGTATTC ATACCACCTG TAGGTAAACA AATTAGAGCA GTAGTTTATT	8220
	ATAATAAAGT AGTTGCATCT AATATGAGTA ATGCTGTTAC TATTTTGCCA GATGACATTC	8280
50	CACCAACAAT CAATAATCCT GTTGAATAA ATGCCAAATA CTATCGAGGC GACGAACAA	8340
	CTTTACAATG GGTGTCTCTG ATAGACATTC TGGTATAAAA AATACAATA TTACGACATT	8400

55

	TACAGGTAGA GTGAGTATGA ATCAGGCATT TAACAGTGAT ATTACATTTA AAGTGTGAGC	8520
	GACAGaCAAT GTCAATAATA CGACAAATGA TAGTCAATCT AAACATGTTT CAATTCATGT	8580
5	AGGTAAATTT AGTGAAGATG CTCATCCGAT TGTATTAGGA AATACTGAGA AAGTTGTAGT	8640
	AGTCAATCCG ACTGCTGTAT CTAATGATGA AAAGCAAAGC ATAATTACTG CCTTTATGAA	8700
	TAAAAACCAA AATATAAGAG GATATTTAGC ATCAACTGAT CCAGTAACTG TCGATAATAA	8760
10	TGGTAATGTC ACATTACATT ACCGTGATGG CTCATCGACA ACGCTTGATG CTACAAATGT	8820
	GATGACATAC GAACCAAGTTG TGAAACCTGA ATACCAAAGT GTCAATGCTG CTAAAACAGC	8880
	AACGGTAACG ATTGCTAAAG GACAATCATT TAGTATTGGT GATATTAAAC AATATTTTAC	8940
15	TTTAAGTAAT GGACAACCTA TTCCAAGTGG CACATTTACA AATATTACAT CTGATAGAAC	9000
	TATTCCAAGT GCACAAGAAG TTAGTCAAAT GAACGCAGGC ACGCAGTTAT ACCATATAAC	9060
	TGCTACAAAT GCGTATCATA AAGATAGTGA AGACTTCTAT ATTAGTTTGA AAATCATCGA	9120
20	TGTGAAACAA CCAGAAGGCG ATCAACGTGT ATATCGTACA TCAACATATG ATTTAACTAC	9180
	TGATGAAATC TCAAAAGTAA AACAAGCATT TATTAATGCA AATAGAGATG TAATTACGCT	9240
25	TGCCGAAGGT GATATTTTCA TTACAAATAC ACCTAATGGT GCTAATGTAA GTACTATTAC	9300
	AGTAAATATT AATAAAGGTC GATTAACGAA ATCATTGCGG TCAAACCTAG CTAATATGAA	9360
	TTTCTTGCCT TGGGTTAATT TCCACAAGA TTATACAGTG ACATGGACGA ATGCAAAAAT	9420
30	TGCAACAGA CCAACAGATG GTGGTTTATC ATGGTCTGAT GACCATAAAT CTTTAATTTA	9480
	TCGTTATGAT GCTACATTAG GTACTCAAAT TACGACGAAT GATATTTTAA CAATGTTAAA	9540
	AGCAACAACT ACAGTGCCTG GATTGCGAAA TAACATTACT GGTAATGAAA AATCACAAGC	9600
35	AGAAGCTGGC GGAAGACCTA ACTTTAGAAC GACTGGTTAT TCACAATCAA ATGCGACAAC	9660
	TGATGGTCAA CGTCAATTTA CGTTGAATGG TCAAGTGATT CAAGTGTTAG ACATCATCAA	9720
	CCCTTCAAAC GGTATGGTG GGCAACCTGT TACAAATTCA AATACTCGTG CAAACCATAG	9780
40	TAACTCAACT GTTGTTAACG TAAACGAACC GGCAGCTAAT GGTGcTGGCG CATTTACAAT	9840
	TGACCACGTT GTAAAAAGTA ATTCTACACA TAATGCAAGT GATGCAGTTT ATAAAGCACA	9900
45	GTTATACTTA ACGCCATATG GTCCAAAACA ATATGTTGAA CATTTAAATC AAAATACAGG	9960
	AAATACTACT GACGCTATTA ACATTTATTT TGTACCAAGT GACTTAGTGA ATCCAACAAT	10020
	TTCAGTAGGT AATTACACTA ATCATCAAGT GTTCTCAGGT GAAACATTTA CAAATACTAT	10080
50	TACAGCGAAT GATAACTTTG GTGTGCAATC TGTAACGTGA CCAAATACAT CACAAATTAC	10140
	AGGTACTGTT GATAATAACC ATCAACATGT TTCTGCAACG GCACCAAATG TGACATCAGC	10200

55

	GTTCAATGTA ACAGTGAAAC CTTTGCCTGA TAAATATCGA GTTGGTACTT CATCAACGGC	10320
	TGCTAATCCT GTGAGAATTG CCAATATTTC GAATAATGCG ACAGTATCAC AAGCTGATCA	10380
5	AACGACAATT ATTAATTCGT TAACGTTTAC TGAAACAGTA CCAAATAGAA GTTATGCAAG	10440
	AGCAAGTGCG AATGAAATCA CTAGTAAAC AGTTAGTAAT GTCAGTCGTA CTGGAAATAA	10500
	TGCCAATGTg cACAGTAACT GTTACTTATC AAGATGGAAC AACATCAACA GTGACTGTAC	10560
10	CTGTAAAGCA TGTCATTCCA GAAATCGTTG CACATTCGCA TTACTCTGTA CAAGGCCAAG	10620
	ACTTCCCAGC AGGTAATGGT TCTAGTGCAT CAGATTACTT TAAGTTATCT AATGGTAGTG	10680
	ACATTGCAGA TGCAACTATT ACATGGGTAA GTGGACAAGC GCCAAATAAA GATAATACAC	10740
15	GTATTGGTGA AGATATAACT GTAAGTGCAC ATATCTTAAT TGATGGCGAA ACAACGCCGA	10800
	TTACGAAAAC AGCAACATAT AAAGTAGTAA GAACTGTACC GAAACATGTC TTTGAAACAG	10860
20	CCAGAGGTGT TTTATACCCA GGTGTTTCAG ATATGTATGA TGCGAAACAA TATGTTAAGC	10920
	CAGTAAATAA TTCTTGGTCG ACAAATGCGC AACATATGAA TTTCCAATTT GTTGAACAT	10980
	ATGGTCCTAA CAAAGATGTT GTAGGCATAT CTACTCGTCT TATTAGAGTG ACATATGATA	11040
25	ATAGACAAAC AGAAGATTTA ACTATTTTAT CTAAAGTTAA ACCTGACCCA CCTAGAATTG	11100
	ACGCAAATC TGAGACATAT AAAGCAGGTC TTACAAACCA AGAAATTAAA GTTAATAACG	11160
	TATTAAATAA CTCGTCAGTA AAATTATTTA AAGCAGATAA TACACCATTA AATGTCACAA	11220
30	ATATTACTCA TGGTAGCGGT TTTAGTTTCG TTGTGACAGT AAGTGACGCG TTACCAAATG	11280
	GCGGAATTAA AGCAAAATCT TCAATTTCAA TGAACAATGT GACGTATACG ACGCAAGACG	11340
	AACATGGTCA AGTTGTTACA GTAACAAGAA ATGAATCTGT TGATTCAAAT GACAGTGCAa	11400
35	CAGTAACAGT GACACCACAA TTACAAGCAA CTACTGAAGG CGCTGTATTT ATTAAAGGTG	11460
	GCGA C GGTTT TGATTTCCGA CACGTAGAAA GATTTATTCA AAACCCGCCA CATGGGGCAA	11520
	CGGT T GCATG GCATGATAGT CCAGATACAT GGAAGAATAC AGTCGGTAAC ACTCATAAAA	11580
40	CTGCGGTTGT AACATTACCT AATGGTCAAG GTACGCGTAA TGTTGAAGTT CCAGTCAAAG	11640
	TTTATCCAGT TGCTAATGCA AAGGCGCCAT CACGTGATGT GAAAGGTCAA AATTTGACTA	11700
45	ATGGAACGGA TGCGATGAAC TACATTACAT TTGATCCAAA TACAAACACA AATGGTATCA	11760
	CTGCAGCATG GGCAAATAGA CAACAACCAA ATAACCAACA AGCAGGCGTG CAACATTTAA	11820
	ATGTCGATGT CACATATCCA GGTATTTTCAG CTGCTAAACG AGTTCCTGTT ACTGTTAATG	11880
50	TATATCAATT TGAATTCCT CAACTACTT ATACGACAAC GGTGGAGGC ACTTTAGCAA	11940
	GTGGTACGCA AGCATCAGGA TATGCACATA TGCAAAATGC TACTGGTTTA CCAACAGATG	12000

55

	TGAATAAACC GAATGTGGCT AAAGTCGTTA ACGCAAATA TGACGTCATC TATAACGGAC	12120
	ATACTTTTGC AACATCTTTA CCAGCGAAAT TTGTAGTAAA AGATGTGCAA CCAGCGAAAC	12180
5	CAACTGTGAC TGAAACAGCG GCAGGAGCGA TTACAATTGC ACCTGGAGCA AACCAAACAG	12240
	TGAATACACA TGCCGGTAAC GTAACGACAT ACGCTGATAA ATTAGTTATT AAACGTAATG	12300
	GTAACGTTGT GACGACATTT ACACGTCGCA ATAATACGAG TCCATGGGTG AAAGAAGCAT	12360
10	CTGCAGCAAC TGTAGCAGGT ATTGCTGGAA CTAATAATGG TATTACTGTT GCAGCAGGTA	12420
	CTTTCAACCC TGCTGATACA ATTCAAGTTG TTGCAACGCA AGGAAGCGGA GAGACAGTGA	12480
	GTGATGAGCA ACGTAGTGAT GATTTACAG TTGTCGCACC ACAACCGAAC CAAGCGACTA	12540
15	CTAAGATTG GCAAAATGGT CATATTGATA TCACGCCTAA TAATCCATCA GGACATTTAA	12600
	TTAATCCAAC TCAAGCAATG GATATTGCTT ACACTGAAAA AGTGGGTAAT GGTGCAGAAC	12660
20	ATAGTAAGAC AATTAATGTT GTTCGTGGTC AAAATAATCA ATGGACAATT GCGAATAAGC	12720
	CTGACTATGT AACGTTAGAT GCACAACTG GTAAAGTGAC GTTCAATGCC AATACTATAA	12780
	AACCAAATTC ATCAATCACA ATTACTCCGA AAGCAGGTAC AGGTCACTCA GTAAGTAGTA	12840
25	ATCCAAGTAC ATTAAGTACA CCGGCAGCTC ATACTGTCAA CACAACTGAA ATTGTGAAAG	12900
	ATTATGGTTC AAATGTAACA GCAGCTGAAA TTAACAATGC AGTTCAAGTT GCTAATAAAC	12960
	GTACTGCAAC GATTAAAAAT GGCACAGCAA TGCCTACTAA TTTAGCTGGT GGTAGCACA	13020
30	CGACGATTCC TGTGACAGTA ACTTACAATG ATGGTAGTAC TGAAGAAGTA CAAGAGTCCA	13080
	TTTTCAAAA AGCGGATAAA CGTGAGTTAA TCACAGCTAA AAATCATTTA GATGATCCAG	13140
	TAAGCACTGA AGGTAAAAAG CCAGGTACAA TTACGCACTA CAATAATGCA ATGCATAATG	13200
35	CGCAACAACA AATCAATACT GCGAAAACAG AAGCACAACA AGTGATTAAAT AATGAGCGTG	13260
	CAACACCACA ACAAGTTTCT GACGCACTAA CTAAAGTTCTG TGCAGCACA ACTAAGATTG	13320
	ATCAAGCTAA AGCATTACTT CAAAATAAAG AAGATAATAG CCAATTAGTA ACGTCTAAAA	13380
40	ATAACTTACA AAGTTCTGTG AACCAAGTAC CATCAACTGC TGGTATGACG CAACAAAGTA	13440
	TTGATAACTA TAATGCGAAG AAGCGTGAAG CAGAACTGA AATAACTGCA GCTCAACGTG	13500
45	TTATTGACAA TGGCGATGCA ACTGCACAAC AAATTTTCTG TGAAAAACAT CGTGTGATA	13560
	ACGCATTAAC AGCATTAAC CAAGCGAAAC ATGATTTAAC TGCAGATACA CATGCCTTAG	13620
	AGCAAGCAGT GCAACAATTG AATCGCACAG GTACAACGAC TGGTAAGAAG CCGGCAAGTA	13680
50	TTACTGCTTA CAATAATTCG ATTCGTGCAC TTCAAAGTGA CTTAACAAGT GCTAAAAATA	13740
	GCGCTAATGC TATTATTCAA AAGCCAATAA GAACAGTACA AGAAGTGCAA TCTGCGTTAA	13800

55

EP 0 786 519 A2

	CTGATAATAG	TGCTTTAAAA	ACTGCTAAGA	CGAAACTTGA	TGAAGAAATC	AATAAATCAG	13920
	TAACTACTGA	TGGTATGACA	CAATCATCAA	TCCAAGCATA	TGAAAATGCT	AAACGTGCGG	13980
5	GTCAAACAGA	ATCAACAAAT	GCACAAAATG	TTATTAACAA	TGGTGATGCG	ACTGACCAAC	14040
	AAATTGCCGC	AGAAAAAACA	AAAGTAGAAG	AAAAATATAA	TAGCTTAAAA	CAAGCAATTG	14100
	CTGGATTAAC	TCCAGACTTG	GCACCATTAC	AAACTGCAAA	AACTCAGTTG	CAAAATGATA	14160
10	TTGATCAGCC	AACGAGTACG	ACTGGTATGA	CAAGCGCATC	TATTGCAGCA	TTTAATGAAA	14220
	AACCTTCAGC	AGCTAGAACT	AAAATTCAAG	AAATTGATCG	TGTATTAGCC	TCACATCCAG	14280
	ATGTTGCGAC	AATACGTCAA	AACGTGACAG	CAGCGAATGC	CGCTAAATCA	GCACTTGATC	14340
15	AAGCACGTAA	TGGCTTAACA	GTCGATAAAG	CGCCTTTAGA	AAATGCGAAA	AATCAACTAC	14400
	AACATAGTAT	TGACACGCAA	ACAAGTACAA	CTGGTATGAC	ACAAGACTCT	ATAAATGCAT	14460
20	ACAATGCGAA	GTTAACAGCT	GCACGTAATA	AGATTCAACA	AATCAATCAA	GTATTAGCAG	14520
	GTTCAACCGAC	TGTAGAACAA	ATTAATACAA	ATACGTCTAC	AGCAAATCAA	GCTAAATCTG	14580
	ATTTAGATCA	TGCACGTCAA	GCTTTAACAC	CAGATAAAGC	GCCGCTTCAA	ACTGCGAAAA	14640
25	CGCAATTAGA	ACAAAGCATT	AATCAACCAA	CGGATACAAC	AGGTATGACG	ACCGCTTCGT	14700
	TAAATGCGTA	CAACCAAAAA	TTACAAGCAG	CGCGTCAAAA	GTAACTGAA	ATTAATCAAG	14760
	TGTTGAATGG	CAACCCAACT	GTCCAAAATA	TCAATGATAA	AGTGACAGAG	GCAAACCAAG	14820
30	CTAAGGATCA	ATTAAATACA	GCACGTCAAG	GTTTAACATT	AGATAGACAG	CCAGCGTTAA	14880
	CAACATTACA	TGGTGCATCT	AACTTAAACC	AAGCACAACA	AAATAATTTT	ACGCAACAAA	14940
	TTAATGCTGC	TCAAATCAT	GctGCGCTTG	AAACAATTAA	GTCTAACATT	ACGGCTTTAA	15000
35	ATACTGCGAT	GACGAAATTA	AAAGACAGTG	TTGCGGATAA	TAATACAATT	AAATCAGATC	15060
	AAAATTACAC	TGACGCAACA	CCAGCTAATA	AACAAGCGTA	TGATAATGCA	GTTAATGCGG	15120
	CTAAAGGTGT	CATTGGAGAA	ACGACTAATC	CAACGATGGA	TGTTAACACA	GTGAACCAAA	15180
40	AAGCAGCATC	TGTTAAATCG	ACGAAAGATG	CTTTAGATGG	TCAACAAAAC	TTACAACGTG	15240
	CGAAAACAGA	AGCAACAAAT	GCGATTACGC	ATGCAAGTGA	TTTAAACCAA	GCACAAAAGA	15300
45	ATGCATTAAC	ACAACAAGTG	AATAGTGcAC	AAAACGTGCA	AGCAGTAAAT	GATATTAAAC	15360
	AAACGACTCA	AAGCTTAAAT	ACTGCTATGA	CAGGTTTAAA	ACGTGGCGTT	GCTAATCATA	15420
	ACCAAGTCGT	ACAAAGTGAT	AATTATGTCA	ACGCAGATAC	TAATAAGAAA	AATGATTACA	15480
50	ACAATGCATA	CAACCATGCG	AATGACATTA	TTAATGGTAA	TGCACAACAT	CCAGTTATAA	15540
	CACCAAGTGA	TGTTAACAAT	GCTTTATCAA	ATGTCACAAG	TAAAGAACAT	GCATTGAATG	15600

55

EP 0 786 519 A2

	ATTTAAATAA TGCACAACGT CAAAACCTTAC AATCGCAAAT TAATGGTGCG CATCAAATTG	15720
	ATGCAGTTAA TACAATTAAG CAAAATGCAA CAACTTGAA TAGTGCAATG GGTAACCTAA	15780
5	GACAAGCTGT TGCAGATAAA GATCAAGTGA AACGTACAGA AGATTATGCG GATGCAGATA	15840
	CAGCTAAACA AAATGCATAT AACAGTGCAG TTTCAAGTGC CGAAACAATC ATTAATCAAA	15900
	CAACAAATCC AACGATGTCT GTTGATGATG TTAATCGTGC AACTTCAGCT GTTACTTCTA	15960
10	ATAAAAATGC ATTAAATGGT TATGAAAAAT TAGCACAATC TAAAACAGAT GCTGCAAGAG	16020
	CAATTGATGC ATTACCACAT TTAAATAATG CACAAAAAGC AGATGTTAA TCTAAAATTA	16080
	ATGCTGCATC AAATATTGCT GCGGTAAATA CTGTTAAACA ACAAGGTACA GATTTAAATA	16140
15	CAKCGATGGg TAACTTGCAA GGTGCAATCA ATGATGAACA AACGACGCTT AATAGTCAAA	16200
	ACTATCAAGA TCGCACACCT AGTAAGAAAA CAGCATACAC AAATGCGGTA CAAGCTGCGA	16260
20	AAGATATTTT AAATAAATCA AATGGTCAAA ATAAAACGAA AGATCAAGTT ACTGAAGCGA	16320
	TGAATCAAGT GAATTCTGCT AAAAAAATACT TAGATGGTAC GCGTTTATTA GATCAAGCGA	16380
	nCAAaCAGCA AAACAGCAGT TAAATAATAT GACGCATTTA ACAACTGCAC AAAAAACGAA	16440
25	TTTAACAAAC CAAATTAATA GTGGTACTAC TGTCGCTGGT GTTCAAACGG TTCAATCAAA	16500
	TGCCAATACA TTAGATCAAG CCATGAATAC GTTAAGACAA AGTATTGCCA ACAAAGATGC	16560
	GACTAAAGCA AGTGAAGATT ACGTAGATGC TAATAATGAT AAGCAAACAG CATATAACAA	16620
30	CGCAGTAGCT GCTGCTGAAA CGATTATTAA TGCTAATAGT AATCCAGAAA TGAATCCAAG	16680
	TACGATTACA CAAAAGCAG AGCAAGTGAA TAGTTCTAAA ACGGCACTTA ACGGTGATGA	16740
	AACTTAGCT GCTGCAAAAC AAAATGCGAA AACGTACTTA AACACATTGA CAAGTATTAC	16800
35	AGATGCTCAA AAGAACAATT TGATTAGTCA AATTACTAGT GCGACAAGAG TGAGTGGTGT	16860
	TGATACTGTA AAACAAAATG CGCAACATCT AGACCAAGCT ATGGCTAGCT TACAGAATGG	16920
40	TATTAACAAC GAATCTCAAG TGAAATCATC TGAGAAATAT CGTGATGCTG ATACAAATAA	16980
	ACAACAAGAG TATGATAATG CTATTACTGC AGCGAAAGCG ATTTTAAATA AATCGACAGG	17040
	TCCAAACACT GCGCAAAATG CAGTTGAAGC AGCATTACAA CGTGTTAATA ATGCGAAAGA	17100
45	TGCATTGAAT GGTGATGCAA AATTAATTGC AGCTCAAAAC GCAGCGAAAC AACATTTAGG	17160
	TACTTTAACG CATATCACTA CAGCTCAACG TAATGATTTA ACAAATCAAA TTTCACAAGC	17220
	TACAACTTA GCTGGTGTG AATCTGTAA ACAAATGCG AATAGTTTAG ATGGTGCTAT	17280
50	GGGTAACCTA CAAACGGCTA TCAACGATAA GTCAGGAACA TTAGCGAGCC AAAACTTCTT	17340
	GGATGCTGAT GAGCAAAAAC GTAATGCATA CAATCAAGCT GTATCAGCAG CCGAAACCAT	17400

55

	TGTTAATAAT GCGAAACATG CATTAAATGG TACGCAAAAC TTAAACAATG CGAAACAAGC	17520
	AGCGATTACA GCAATCAATG GCGCATCTGA TTTAAATCAA AAACAAAAAG ATGCATTAAA	17580
5	AGCACAAGCT AATGGTGCTC AACGCGTATC TAATGCACAA GATGTACAGC ACAATGCGAC	17640
	TGAACTGAAC ACGGCAATGG GCACATTAAA ACATGCCATC GCAGATAAGA CGAATACGTT	17700
	AGCAAGCAGT AAATATGTTA ATGCCGATAG CACTAAACAA AATGCTTACA CAACTAAAGT	17760
10	TACCAATGCT GAACATATTA TTAGCGGTAC GCCAACGGTT GTTACGACAC CTTCAGAAGT	17820
	AACAGCTGCA GCTAATCAAG TAAACAGCGC GAAACAAGAA TTAAATGGTG ACGAAAGATT	17880
	ACGTGAAGCA AAACAAAACG CCAATACTGC TATTGATGCA TTAACACAAT TAAATACACC	17940
15	TCAAAAAGCT AAATTAAAAG AACAAAGTGGG ACAAGCCAAT AGATTAGAAG ACGTACAAAC	18000
	TGTTCAAACA AATGGACAAG CATTGAACAA TGCAATGAAA GGCTTAAGAG ATAGTATTGC	18060
20	TAACGAAACA ACAGTCAAAA CAAGTCAAAA CTATACAGAC GCAAGTCCGA ATAACCAATC	18120
	AACATATAAT AGCGCTGTGT CAAATGCGAA AGGTATCATT AATCAAATA ACAATCCGAC	18180
	TATGGATACT AGTGCGATTA CCCAAGCTAC AACACAAGTG AATAATGCTA AAAATGGTTT	18240
25	AAACGGTGCT GAAAACCTTA GAAATGCACA AAACACTGCT AAGCAAACT TAAATACATT	18300
	ATCACACTTA ACAAATAACC AAAAATCTGC CATCTCATCA CAAATTGATC GTGCAGGTCA	18360
	TGTGAGTGAG GTAACGCTA CTAAAAATGC AGCAACTGAG TTGAATACGC AAATGGGTAA	18420
30	CTTGGAACAA GCTATCCATG ATCAAAACAC AGTTAAACAA AGTGTTAAAT TTAATGATGC	18480
	AGATAAAGCT AAACGTGATG CGTATACAAA TGCGGTAAGC AGAGCTGAAG CAATTCTGAA	18540
	TAAAACGCAA GGTGCAATA CGTCTAAACA AGATGTTGAA GCGGCTATTC AAAATGTTTC	18600
35	AAGTGCTAAA AATGCATTGA ATGGTGATCA AAACGTTACA AATGCGAAGA ATGCAGCTAA	18660
	AAATGCATTA AATAACTTAA CGTCAATTAA TAATGCACAA AAACGTGACT TAACAACTAA	18720
	AATTGATCAA GCAACAACTG TAGCTGGTGT TGAAGCTGTA TCTAATACGA GTACACAATT	18780
40	GAATACAGCG ATGGCTAACT TGCAAAATGG TATTAATGAT AAAACAAATA CACTAGCAAG	18840
	TGAAAACCTAT CATGATGCTG ATTCAGATAA GAAAACCTGCT TATACTCAAG CCGTTACGAA	18900
45	CGCAGAAAAT ATTTTAAATA AAAATAGTGG ATCAAATTTA GACAAAACCTG CCGTTGAAAA	18960
	CGCGTTGTCA CAAGTTGCTA ATGCGAAAGG TGCCCTAAAT GGTAACCATA ATTTAGAGCA	19020
	AGCTAAATCA AATGCAACA CTACTATAAA CGGACTTCAA CATTTAACAA CTGCTCAAAA	19080
50	AGATAAATTG AAACAACAAG TGCAACAAGC ACAAATGTT GCAGGTGTAG ATACTGTTAA	19140
	ATCAAGTGCC AACACATTAA ATGGTGCTAT GGGTACGTTA AGAAATAGCA TACAAGATAA	19200

55

	TAACAATGCT GTTGATAGTG CTAATGGTGT CATTAAATGCA ACAAGCAATC CAAATATGGA	19320
	TGCTAATGCA ATTAACCAAA TCGCTACACA AGTGACATCA ACGAAAAATG CATTAGATGG	19380
5	TACACATAAT TTAACGCAAG CGAAACAAAC AGCAACAAAT GCCATCGATG GTGCTACTAA	19440
	CTTAAATAAA GCGCAAAAAG ATGCGTTAAA AGCACAAAGTT ACAAGTGCGC AACGTGTTGC	19500
	AAATGTAACA AGTATCCAAC AAAGTCAAA TGAAGTTAAT ACAGCTATGG GTCAATTACA	19560
10	ACATGGTATT GATGATGAAA ATGCAACAAA ACAAGTCAA AAATATCGTG ACGcTGAACA	19620
	AAGTAAGAAA ACTGCTTATG ATCAAGCTGT AGCTGCTGCG AAAGCAATTT TAAATAAACA	19680
	AACAGGTTCA AATTCAGATA AAGCAGCAGT TGACCGTGCA TTACAACAAG TAACAAGTAC	19740
15	GAAAGATGCA TTGAATGGTG ATGCAAAACT GGCAGAAGCG AAAGCGGCAG CTAAACAAAA	19800
	CTTAGGCACT TTAAACCATA TTACGAATGC ACAACGTACT GACTTAGAAG GCCAAATCAA	19860
	TCAAGCGACG ACTGTTGATG GCGTTAATAC TGTAATAACA AATGCCAATA CATTAGACGG	19920
20	CGCAATGAAT AGCTTACAAG GTTCAATCAA TGATAAAGAT GCGACATTAA GAAATCAAAA	19980
	TTATCTTGAT GCGGATGAAT CAAAACGAAA TGCATATACG CAAGCTGTCA CAGCGGCTGA	20040
25	AGGCATTTTA AATAAACAAA CTGGTGGTAA CACATCTAAA GCAGACGTTG ATAATGCATT	20100
	AAATGCAGTT ACAAGAGCGA AAGcGgCTTT AAATGGTGCT GACAAGTTAA GAAATGCGAA	20160
	AACTTCAGCA ACAATACGA TTGATGGTTT ACCTAACTTA ACACAATTAC AAAAAGACAA	20220
30	CTTGAAGCAT CAAGTTGAaC AAGCGCAAAA TGTAGCAGGT GTAAATGGTG TTAAAGATAA	20280
	AGGTAATACG TTAAATACTG CCATGGGTGC ATTACGTACA AGTATCCAAA ATGATAATAC	20340
	GACGAAAACA AGTCAAAATT ATCTTGATGC ATCTGACAGC AACAAAAATA ATTACAATAC	20400
35	TGCTGTAAAT AATGCAAATG GTGTTATTAA TGCAACGAAC AATCCAAATA TGGATGCTAA	20460
	TGCGATTAAAT GGCATGGCAA ATCAAGTCAA TACAACAAAA GCAGCGTTAA ATGGTGCACA	20520
	AAACTTAGCT CAAGCTAAAA CAAATGCGAC GAACACAATT AACAAACGCAC ATGACTTAAA	20580
40	CCAAAAACAA AAAGATGCAT TAAAAACACA AGTTAACAAT GCACAACGTG TATcTGATGC	20640
	AAATAACGTT CAACACACTG CAACTGAATT GAACAGTGCG ATGACAGCAC TTAAAGCAGC	20700
45	TATTGCTGAT AAAGAAAGAA CAAAAGCAAG CGGTAATTAT GTCAATGCTG ATCAAGAAAA	20760
	ACGTCAAGCG TATGATTCAA AAGTGACTAA CGCTGAAAAT ATCATTAGTG GTACACCGAA	20820
	TGCGACATTA ACAGTCAATG ACGTAAATAG TGCGGCATCA CAAGTCAATG CGGCTAAAC	20880
50	AGCATTAAAT GGTGATAACA ACTTACGTGT AGCGAAAGAG CATGCCAACA ATACAATTGA	20940
	CGGCTTAGCA CAATTGAATA ATGCACAAAA AGCAAAATTA AAAGAACAAG TTCAAAGTGC	21000

55

	GAAAGGCTTA AGAGATAGTA TTGCGAATGA AGCAACAATT AAAGCAGGTC AAAACTACAC	21120
	TGACGCAAGT CCAAATAATC GTAACGAGTA CGACAGTGCA GTTACTGCAG CAAAAGCAAT	21180
5	CATTAATCAA ACATCGAACC CAACGATGGA ACCAAATACT ATTACGCAAG TAACATCACA	21240
	AGTGACAACCT AAAGAACAGG CATTAAATGG TGCGCGAAAC TTAGCTCAAG CTAAGACAAC	21300
	TGCGAAAAAC AACTTGAATA ACTTAACATC AATTAACAAT GCACAAAAAG ATGCGTTAAC	21360
10	GCGTAGcATT GATGGTGCAA CAACAGTAGC TGGTGTAAT CAAGAACTG CAAAAGCAAC	21420
	AGAATTAAAT AACGCAATGC ATAGTTTACA AAATGGTATC AATGATGAGA CACAAACAAA	21480
	ACAAACTCAG AAATACCTAG ATGCAGAGCC AAGTAAGAAA TCAGCTTATG ATCAAGCAGT	21540
15	AAATGCAGCG AAAGCAATTT TAACAAAAGC TAGTGGTCAA AATGTAGACA AAGCAGCAGT	21600
	TGAACAAGCA TTGCAAAATG TGAACAGTAC GAAGACGGCG TTGAACGGTG ATGCGAAATT	21660
20	AAATGAAGCT AAAGCAGCTG CGAAACAAAC GTTAGGTACA TTAACACACA TTAATAATGC	21720
	ACAACGTACA GCGTTAGACA ATGAAATTAC ACAAGCAACA AATGTTGAAG GTGTTAATAC	21780
	AGTTAAAGCC AAAGCGCAAC AATTAGATGG TGCTATGGGT CAATTAGAAA CATCAATTCG	21840
25	TGATAAAGAC ACGACGTTAC AAAGTCAAAA TTATCAAGAT GCTGATGATG CTAAACGAAC	21900
	TGCTTATTCT CAAGCAGTAA ATGCAGCAGC AACTATTTTA AATAAAACag CTGGCGGTAA	21960
	TACACCTAAA GCAGATGTTG AAAGAGCAAT GCAAGCTGTT ACACAAGCAA ATACTGcATT	22020
30	AAACGGTATT CAmAaCTTAG ATCGTGCGAA ACaTgCTGCT AACACAGCGA TTACAAATGC	22080
	TTCGGACTTA AATACAAAAC mAAAAGAAGC ATTAAAagCA CAAGTAACAA GTGCAGGACG	22140
	TGTATCTGCA GCAAATGGTG TTGAACATAC TGCGACTGAA TTAAATACTG CGATGACAGC	22200
35	TTTAAAGCGT GCCATTGCTG ATAAAGCTGA GACAAAAGCT AGTGGTAACT ATGTCAATGC	22260
	TGATGCGAAT AAACGTCAAG CATATGATGA AAAAGTTACA GCTGCCGAAA ATATCGTTAG	22320
	TGGTÀCACCÀ ACACCAACGT TAACACCAGC AGATGTTACA AATGCAGCAA CGCAAGTAAC	22380
40	GAATGCTAAG ACGCAGTTAA ACGGTAATCA TAATTTAGAA GTAGCGAAAC AAAATGCTAA	22440
	CACTGCAATT GATGGTTTAA CTTCTTTAAA TGGTCCGCAA AAAGCAAAAC TTAAAGAACA	22500
45	AGTGGGTCAA GCGACGACGT TGCCAAATGT TCAAAGTGT CGTGATAATG CACAAACATT	22560
	AAACACTGCA ATGAAAGGTC TACGAGATAG CATTGCGAAT GAAGCAACGA TTAAAGCAGG	22620
	TCAAACTAC ACAGATGCAA GTCAAAACAA ACAAACTGAC TACAACAGTG CAGTCACTGC	22680
50	AGCAAAAGCA ATCATTGGTC AAACAAGTAG TCCATCAATG AATGCGCAAG AAATTAATCA	22740
	AGCGAAAGAC CAAAGTACAG CTAAACAACA AGCGTTAAAC GGTCAAGAAA ACTTAAGAAC	22800

55

	AGATGCAGTG AAACGTCAAA TCGAAGGTGC AACGCATGTT AATGAAGTAA CACAAGCACA	22920
	AAATAATGCG GATGCATTAA ATACAGCTAT GACGAAGTTG AAAAATGGTA TTCAAGATCA	22980
5	GAATACGATT AAGCAAGGTG TTAACCTTCAC TGATGCCGAC GAAGCGAAAC GTAATGCATA	23040
	TACAAATGCA GTGACGCAAG CTGAACAAAT TTTAAATAAA GCACAAGGTC CAAATACTTC	23100
	AAAAGACGGT GTCGAAACTG CGTTAGAAaAA TGTACAACGT GCTAAAAACG AATTGAACGG	23160
10	TAATCAAAAT GTTGCGAACG CTAAGACAAC TGCGAAAAAT GCATTGAATA ACCTAACATC	23220
	AATTAATAAT GCACAAAAAG AAGCATTGAA ATCACAAATT GAAGGTGCGA CAACAGTTGC	23280
	AGGTGTAAAT CAAGTGTCTA CAACGGCATC TGAATTAAAT ACAGCAATGA GCAACTTACA	23340
15	AAATGGTATT AATGATGAAG CAGCTACAAA AGCAGCGCTT AATGGTACTC AAAACCTTGA	23400
	AAAAGCTAAA CAACACGCAA ATACAGCAAT TGACGGTTTA AGCCATTTAA CAAATGCACA	23460
20	AAAAGAGGCA TTAAACAAT TGGTACAACA ATCGACTACT GTTGCGAAG CACAAGGTAA	23520
	TGAGCAAAAA GCAACAATG TTGATGCAGC AATGGACAAA TTACGTCAAA GTATTGCAGA	23580
	TAATGCGACA ACAAACAAA ACCAAAATTA TACTGATGCA AGTCAGAATA AAAAGGATGC	23640
25	GTACAATAAT GCTGTCACAA CTGCACAAGG TATTATTGAT CAAACTACAA GTCCAACTTT	23700
	AGATCCGACT GTTATCAATC AAGCTGCTGG ACAAGTAAGC ACAACTAAAA ATGCATTAAA	23760
	TGGTAATGAA AACCTAGAGG CAGCGAAACA ACAAGCGTCA CAATCATTAG GTTCATTAGA	23820
30	TAACTTAAAT AATGCGCAAA AACAAACAGT TACTGATCAA ATTAATGGCG CGCATACTGT	23880
	TGATGAAGCA AATCAAATTA AGCAAAATGC GCAAACTTA AATACAGCGA TGGGTAACTT	23940
	GAAACAAGCG ATAGcTGACA AAGATGCTAC GAAAGCGACA GTTAACTTCA CTGATGCAGA	24000
35	TCAAGCAAAA CAACAAGCAT ATAACaCTGC TGTTACAAAT GCTGAAAATA TCATTTCAAA	24060
	AGCTAATGGC GGCAATGCAA CACAAGCTGA AGTTGAACAA GCAATCAAAC AAGTTAATGC	24120
40	TGCAAAACAA GCATTAAATG GTAATGCCAA CGTTCAACAT GCAAAAGACG AAGCAACAGC	24180
	ATTAATTAAT AGCTCTAATG ACCTTAACCA AGCACAAAAA GACGCATTAA AACAACAAGT	24240
	TCAAAATGCA ACTACTGTAG CTGGTGTAAT CAATGTTAAA CAAACAGCAC AAGAGTTAAA	24300
45	CAATGCTATG ACACAATTAA AACAAGGCAT TGCAGATAAA GAACAAACAA AAGCTGATGG	24360
	TAACCTTGTC AATGCAGATC CTGATAAGCA AAATGCATAT AATCAAGCAG TAGCGAAAGC	24420
	TGAAGCATTa ATTAGTGctA CGCCTGATGT TGTGTTACA CCTAGCGAAA TTACTGCAGC	24480
50	GTAAATAAAA GTTACGCAAG CTAAAAATGA TTAAATGGT AATACAAACT TAGCAACGGC	24540
	GAAACAAAAT GTTCAACATG CTATTGATCA ATTGCCAAAC TTAAACCAAG CGCAACGTGA	24600

55

	AGCGGCGACA ACGCTTAATG ACGCGATGAC ACAATTGAAA CAAGGTATTG CGAATAAAGC	24720
	ACAAATTAAA GGTAGCGAGA ACTATCACGA TGCTGATACT GACAAGCAAA CAGCATATGA	24780
5	TAATGCAGTA ACAAAGCAG AAGAATTGTT AAAACAAACA ACAAATCCAA CAATGGATCC	24840
	AAATACAATT CAACAAGCAT TAACTAAAGT GAATGACACA AATCAAGCAC TTAACGGTAA	24900
	TCAAAAATTA GCTGATGCCA AACAAGATGC TAAGACAACA CTTGGTACAC TAGATCATTT	24960
10	AAATGATGCT CAAAAACAAG CGCTAACAAAC TCAAGTTGAA CAAGCACCAG ATATTGCAAC	25020
	AGTTAATAAT GTTAAGCAAA ATGCTCAAAA TCTGAATAAT GCTATGACTA ACTTAAACAA	25080
	TGCATTACAA GATAAACTG AGACATTAAA TAGCATTAAAC TTTACTGATG CAGATCAAGC	25140
15	TAAGAAAGAT GCTTATACTA ATGCGGTTTC ACATGCAGAA GGTATTTTAT CTAAAGCAAA	25200
	TGGCAGCAAT GCAAGTCAAA CTGAAGTGA ACAAGCGATG CAACGTGTGA ACGAAGCGAA	25260
20	ACAAGCATTG AATGGTAATG ACAATGTACA ACGTGCAAAA GATGCAGCGA AACAAGTGAT	25320
	TACAAATGCA AATGATTTAA ATCAAGCAAT GACACAATTG AAACAAGGTA TTGCAGATAA	25380
	AGACCAAACCT AAAGCAAATG GTAACCTTGT CAATGCTGAT ACTGATAAGC AAAATGCTTA	25440
25	CAACAATGCG GTAGCACATG CTGAACAAAT AATTAGTGGT ACACCAAATG CAAACGTGGA	25500
	TCCACAACAA GTGGCTCAAG CGTTACAACA AGTGAATCaA GCTAAGGGTG ATTTAAACGG	25560
	TAACCATAAC TTACAAGTTG CTAAAGACAA TGCAAATACA GCCATTGATC AGTTACCAAA	25620
30	CTTAAATCAA CCACAAAAAA CAGCATTAAA AGACCAAGTG TCGCATGCAG AACTTGTTAC	25680
	AGGTGTTAAT GCTATTAAGC AAAATGCTGA TCGGTTAAAT AATGCaATGG GTACATTGAA	25740
	ACAACAAATT CAAGCGAACA GTCAAGTACC ACAGTCAGTT GACTTTACAC AAGCGGATCA	25800
35	AGACAAACAA CAAGCATATA ACAATGCGGC TAACCAAGCG CAACAAATCG CAAATGGCAT	25860
	ACCAACACCT GTATTGACGC CTGATACAGT AACACAAGCA GTGACAACTA TGAATCAAGC	25920
	GAAAGATGCA TTAAACGGTG ATGAAAAATT AGCACAAGCG AAACAAGAAG CTTTAGCAAA	25980
40	TCTTGATACG TTACGCGATT TAAATCAACC ACAACGTGAT GCATTACGTA ACCAAATCAA	26040
	TCAAGCACAA GCGTTAGCTA CAGTTGAACA AACTAAACAA AATGCACAAA ATGTGAATAC	26100
45	aGCaATGAGT AACTTGAAAC aAGGTATTGC aAACAAAGAT ACTGTCAAAG CAAGTGAGAA	26160
	CTATCATGAT GCTGATGCCG ATAAGCAAAC AGCATATACA AATGCAGTGT CTCAAGCGGA	26220
	AGGTATTATC AATCAAACGA CAAATCCAAC GCTTAACCCA GATGAAATAA CACGTGCATT	26280
50	AACTCAAGTG ACTGATGCTA AAAATGGCTT AAACGGTGAA GCTAAATTGG CAACTGAAAA	26340
	GCAAAATGCT AAAGATGCCG TAAGTGGGAT GACGCATTTA AACGATGCTC AAAACAAGC	26400

55

	AGCAACGAGC CTAGATCAAG CAATGGATCA ATTATCACAA GCTATTAATG ATAAAGCTCA	26520
	AACATTAGCG GACGTAATT ACTTAAATGC AGATCCTGAC AAACAAAATG CGTATAAACA	26580
5	GGCAGTAGCA AAAGCTGAAG CATTATTGAA TAAACAAAGT GGTACTAATG AAGTACAAGC	26640
	ACAAGTTGAA AGCATCACTA ATGAAGTGAA CGCAGCGAAA CAAGCATTAA ATGGTAATGA	26700
	CAATTTGGCA AATGCAAAAC AACAAGCAAA ACAACAATTG GCGAACTTAA CACACTTAAA	26760
10	TGATGCACAA AAACAATCAT TTGAAAGTCA AATTACACAA GCGCCACTTG TTACAGATGT	26820
	CACTACGATT AATCAAAAAG CACAAACGTT AGATCATGCG ATGGAATTAT TAAGAAATAG	26880
	TGTTGCGGAT AATCAAACGA CATTAGCGTC TGAAGATTAT CATGATGCAA CTGCGCAAAG	26940
15	ACAAAATGAC TATAACCAAG CTGTAACAGC TGCTAATAAT ATAATTAATC AAACCTACATC	27000
	GCCTACGATG AATCCAGATG ATGTTAATGG TGCAACGACA CAAGTGAATA ATACGAAAGT	27060
20	TGCATTAGAT GGTGATGAAA ACCTTGCGAGC AGCTAAACAA CAAGCAAACA ACAGACTTGA	27120
	TCAATTAGAT CATTTGAATA ATGCGCAAAA GCAACAGTTA CAATCACAAA TTACGCAATC	27180
	ATCTGATATT GCTGCAGTTA ATGGTCACAA ACAACAGCA GAATCTTTAA ATACTGCGAT	27240
25	GGGTAACCTA ATTAATGCGA TTGCAGATCA TCAAGCCGTT GAACAACGTG GTAACCTCAT	27300
	CAATGCTGAT ACTGATAAAC AAACCTGCTTA TAATACAGCG GTAAATGAAG CAGCAGCAAT	27360
	GATTAACAAA CAAACTGGTC AAAATGCGAA CCAAACAGAA GTAGAACAAG CTATTACTAA	27420
30	AGTTCAAACA ACACCTCAAG CGTTAAATGG AGACCATAAT TTACAAGTTG CTAAAACAAA	27480
	TGCGACGCAA GCAATTGATG CTTTAACAAG CTTAAATGAT CCTCAAAAAA CAGCATTAAA	27540
	AGACCAAGTT ACAGCTGCAA CTTTAGTAAC TGCAGTTCAT CAAATTGAAC AAAATGCGAA	27600
35	TACGCTTAAC CAAGCAATGC ATGGTTTAAAG ACAGAGCATT CAAGATAACG CAGCAACTAA	27660
	AGCAATATAGC AAATATATCA ACGAAGATCA ACCAGAGCAA CAAACTATG ATCAAGCTGT	27720
40	TCAAGCCGCA AATAATATTA TCAATGAACA AACTGCAACA TTAGATAATA ATGCGATTAA	27780
	TCAAGCAGCG ACAACTGTGA ATACAACGAA AGCAGCATT CATGGTGATG TGAAGTTACA	27840
	AAATGATAAA GATCATGCTA AGCAAACGGT TAGTCAATTA GCACATCTAA ACAATGCACA	27900
45	AAAACATATG GAAGATACGT TAATTGATAG TGAAACAACT AGAACAGCAG TTAAGCAAGA	27960
	TTTGACTGAA GCACAAGCAT TAGATCAACT TATGGATGCA TTACAACAAA GTATTGCTGA	28020
	CAAAGATGCA ACACGTGCGA GCAGTGCATA TGTCAATGCA GAACCGAATA AAAAACAATC	28080
50	CTATGATGAA GCAGTTCAAA ATGCTGAGTC TATCATTGCA GGATTAAATA ATCCAATAT	28140
	CAATAAAGGT AATGTATCAA GTGCGACTCA AGCAGTAATA TCATCTAAAA ATGCATTAGA	28200

55

	TCAATTAACA CCAGCTCAAC AACAAAGCGCT AGAAAATCAA ATTAATAATG CAACAACCTCG	28320
	TGATAAAGTG GCTGAAATCA TTGCACAAGC GCAAGCATtA AATGAAGCGA TGAAAGCATT	28380
5	AAAAGAAAGT ATTAAGGATC AACCACAAAC TGAAGCAAGT AGTAAATTTA TTAACGAGGA	28440
	TCAAGCGCAA AAAGATGCTT ATACGCAAGC AGTACAACAC GCGAAAGATT TGATTAACAA	28500
	AACAACCTGAT CCTACATTAG CTAAATCAAT CATTGATCAA GCGACACAGG CAGTGACAGA	28560
10	TGCTAAAAAC AATTTACATG GTGATCAAAA ACTAGCTCAA GATAAGCAAC GTGCAACAGA	28620
	AACGTTAAAT AACTTGTCTA ACTTGAATAC ACCACAACGT CAAGCACTTG AAAATCAAAT	28680
	TAATAATGCA GCAACTCGTG GCGAAGTAGC ACAAAAATTA ACTGAAGCAC AAGCACTTAA	28740
15	CCAAGCAATG GAAGCTTTAC GTAATAGCAT TCAAGATCAA CAGCAAACGG AAGCGGGTAG	28800
	CAAGTTTATC AATGAAGATA AaCCaCmAAA AGrTGCTTAC CAAGCAGCAG TTCAAAATGC	28860
20	AAAAGATTTA ATTAATCAAA CTAACAATCC AACGCTTGAT AAAGCACAAG TTGAACAATT	28920
	GACACAAGCT GTTAACCAAG CTAAAGATAA CCTACACGGT GATCAAAAAC TTGCAGACGA	28980
	TAAACAACAT GCGGTTACTG ATTTAAATCA ATTAAATGGT TTGAATAATC CGCAACGTCA	29040
25	AGCACTTGAA AGCCAAATAA ACAACGCAGC AACTCGTGGC GAAGTAGCAC AAAAATTAGC	29100
	TGAAGCAAAA GCGCTTGATC AAGCAATGCA AGCATTACGT AATAGTATTC AAGATCAACA	29160
	ACAAACAGAA TCTGGTAGCA AGTTTATCAA TGAAGATAAA CCGCAAAAAG ATGCTTACCA	29220
30	AGCAGCAGTT CAAATGCAA AAGATTTAAT TAACCAACA GGTAATCCAA CACTCGACAA	29280
	ATCACAAGTA GAACAATTGA CACAAGCAGT AACAACTGCA AAAGATAATC TACATGGTGA	29340
	TCAAAAACCTT GCTCGTGATC AACACAAGC AGTAACAACT GTAAATGCAT TGCCAACTT	29400
35	AAATCATGCA CAACAACAAG CATTAACTGA TGCTATAAAT GCAGCGCCTA CAAGAACAGA	29460
	GGTTGACAA CATGTTCAA CTGCTACTGA ACTTGATCAC GCGATGGAAA CATTGAAAAA	29520
	TAAAGTTGAT CAAGTGAATA CAGATAAGGC TCAACCAAAT TACACTGAAG CGTCAACTGA	29580
40	TAAAAAAGAA GCAGTAGATC AAGCGTTACA AGCTGCAGAA AGCATTACAG ATCCAACATA	29640
	TGGTTCAAAT GCGAATAAAG ACGCTGTAGA CCAAGTATTA ACTAAGCTTC AAGAAAAAGA	29700
45	AAATGAGTTA AATGGTAATG AGAGAGTCGC TGAAGCTAAA ACACAAGCGA AACAACTAT	29760
	TGACCAATTA ACACATTTAA ATGCTGATCA AATTGCAACT GCTAAACAAA ACATTGATCA	29820
	AGCGACGAAA CTTCAACCAA TTGCTGAATT AGTAGATCAA GCAACGCAAT TGAATCAATC	29880
50	TATGGATCAA TTACAACAAG CAGTTAATGA ACATGCTAAC GTTGAGCAAA CTGTAGATTA	29940
	CACACAAGCA GATTCAGATA AACAAAATGC TTATAAACAA GCTATTGCTG ATGCTGAAAA	30000

55

EP 0 786 519 A2

	TGCAAAACAA GCATTAAATG GTGATGAACG TGTAGCACTT GCTAAAACAA ATGGTAAACA	30120
	TGACATCGAC CAATTGAATG CATTAAACAA TGCTCAACAA GATGGATTTA AAGGTCGCAT	30180
5	CGATCAATCA AACGATTTAA ATCAAATCCA ACAAATTGTA GATGAGGCTA AGGCACTTAA	30240
	TCGTGCAATG GATCAATTGT CACAAGAAAT CACTGACAAT GAAGGACGCA CGAAAGGTAG	30300
	CACGAACTAT GTCAATGCAG ATACACAAGT CAAACAAGTA TATGATGAAA CGGTTGATAA	30360
10	AGCGAAACAA GCACTTGATA AATCGACTGG TCAAACTTA ACTGCAAAAC AAGTTATCAA	30420
	ATTAAATGAT GCAGTCACTG CAGCTAAGAA AGCATTAAAT GGTGAAGAAA GACTTAATAA	30480
	TCGTAAAGCT GAAGCATTAC AAAGATTGGA TCAATTAACA CATCTAAACA ATGCTCAAAG	30540
15	ACAATTAGCA ATCCAACAAA TTAATAATGC TGAAACGCTA AATAAAGCAT CTCGAGCAAT	30600
	TAATAGAGCA ACTAAATTAG ATAATGCAAT GGGTTCAGTA CAACAATATA TTGACGAACA	30660
20	GCACCTTGGT GTTATCAGCA GCACAAATTA CATCAATGCA GATGACAATT TGAAAGCAAA	30720
	TTATGATAAT GCAATTGCGA ATGCAGCACA TGAGTTAGAT AAAGTGCAAG GTAATGCAAT	30780
	TGCcAAAAGCT GAAGCAGAGC AATTGAAACA AAATATTATC GATGCTCAA ATGCATTAAA	30840
25	TGGAGACCAA AACCTTGCAA ATGCCAAAGA TAAAGCAAAT GCGTTTGTTA ATTCGTTAAA	30900
	TGGATTAAAT CAACAGCAAC AAGATCTTGC ACATAAAGCA ATTAACAATG CCGATACTGT	30960
	ATCAGATGTA ACAGATATTG TTAATAATCA AATTGACTTA AATGATGCAA TGGAAACATT	31020
30	GAAACATTTA GTTGACAATG AAATTCCAAA TGCAGAGCAA ACTGTCAATT ACCAAAACGC	31080
	TGACGATAAT GCTAAA	31096

(2) INFORMATION FOR SEQ ID NO: 60:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 2243 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
40	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

45	ATGACAGAAT GGGAGCGAGG ACTTAGAATG TTTCCTAAAT CAGGTTTATT AAATTTTGAG	60
	TTAGCGATAG mAAATCGTTC ATTAAATGAT GATGAAAAAG CATTAAAATA TGTGCGTAAA	120
	GCATTAAATG CAGACCCTAA AAATACAGAT TATATTAAC TAGAAAAAGA GTTGACTAAA	180
50	TCAATGAGT CGAAAAATAA ATAACTTTTA TGATGTACAA CAGTTATTGA AAAGTTACGG	240
	ATTTCTAATA TATTTTAAAA ATCCAGAAGA TATGTACGAA ATGATTCAAC AGGAGATTTC	300

55

	TAATCAGAGA AGGAATGAAC AGAAATGACA AAAATTATTT TAGCAGCTGA TGTAGGCGGG	420
	ACGACTTGTA AATTAGGTAT TTTCACACCT GAATTAGAAC AATTACATAA ATGGTCTATT	480
5	CACACTGATA CATCTGATAG TACAGGATAT ACACTTTTGA AAGGAATTTA TGATTCGTTT	540
	GTTGAAAAAG TAAATGAAAA TAATTATAAT TTTTCAAATG TACTTGCGCT AGGTATTGGT	600
	GTACCAGGTC CTGTTGACTT TGAAAAAGGT ACAGTAAATG GAGCAGTAAA CTTATATTGG	660
10	CCAGAAAAAG TTAATGTACG TGAGATTTTT GAACAATTCG TTGATTGTCC AGTGTATGTA	720
	GATAATGATG CTAACATAGC TGCTTTAGGG GaGAAACACA AAGGTGCTGG TGAAGGTGCC	780
15	GATGATGTTG TTGCCATCAC ACTTGGTACA GGTCTAGGTG GAGGAATTAT TTCCAAATGG	840
	TGAAATCGTA CATGGTCATA ATGGCTCtGG CGCAGAAATA GGTCATTTTA GAgCAGACTT	900
	CgATCAACGA TTTaAATGTA ATTGTGGTCG TTCTGGATGT ATTGAAACAG TTGCTTCaGC	960
20	GACAGGCGTT GTTAACTTAG TTAACTTcLa CTATCCGAAG TTGACGTTTA GATCTTCTAT	1020
	ATTAGAATTG ATTAAAGAAA ATAAGGtTAC aGCAAAAGCT GTTTTTGATG CGGCAAAAGC	1080
	TGGTGACCAA TTCTGTATTT TCATTACTGA AAAGGTTGCA AACTATATTG GATATTTATG	1140
25	TAGTATTATT AGTGTTACAA GTAATCCGAA ATATATCGTT CTAGGTGGAG GAATGTCTAC	1200
	TGCAGGACCT ATTTTAATTG AAAATATTAA AACAGAATAT CATAATTTAA CATTTCACC	1260
	TGCTCAATTT GAAACTGAAA TTGTACAAGC GAAATTAGGT AATGATGCAG GTATTACAGG	1320
30	AGCAGCAGGA TTAATCAAGA CCTATGTATT AGATAAAGAG GGGGTAAAAT AATGGCTATT	1380
	GTTGATGTGG TTGTTATTCC AGTTGGAACG GAAGGTCCGA GTGTTAGTAA ATATATTGCA	1440
35	GATATTCAGA AAAAaCTTCA AGAATATAAA GCAATGGGTA AAATTGATTT TCAATTAACA	1500
	CCAATGAATA CTCTAATTGA AGGTGAATTA AGCGATGTAT TAGAAGTTGT GCAAGTGATA	1560
	CATG A ATTAC CTTTtGATAA AGGTTTAAGT AGAGTTTGTa CAAATATCCG TATTGATGAC	1620
40	CGACGAGACA AATCTAGAAA AATGAATGAT AAACTAACAT CAGTACAAAA ACATTTAGAA	1680
	AATAGTGGTG AAAACCTATG AGGATTTCAA GCTTAACTTT AGGCTTAGTT GATACTAATA	1740
	CGTATTTcAT CGAAAATGAC AAAGCTGTTA TTCTGATTGA CCCTTCAGGT GAAAGTGAAA	1800
45	AAATTATTAA AAAATTAAAC CAAATAAATA AACCGTTAAA AGCTATTTTA TTAACACATG	1860
	CACACTTTGA TCATATCGGA GCAGTCGATG ATATAGTTGA TCGATTGAT GTCCCGGTTT	1920
	ATATGCATGA AGCAGAGTTT GATTTTCTAA AAGATCCCGT TAAAAATGGG GCAGATAAAT	1980
50	TTAAGCAATA TGGATTACCA ATTATTACAA GTAAGGTAAC TCCTGAAAAG TTAaMCGAAG	2040
	GTAGCACAGA AATAGAAGGA TTTAAGTTnT nAyrTGtCa CACACCTGGA CATTcACCAG	2100

55

GAATCGGACG TACAGATTTA TATAAAGGTG ATTATGAAAC GCTAGTTGAT TCTATTCAAG 2220
 ATAAAATATT TGAATTAGAA GGC 2243

5 (2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8009 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

15 TTGGnATCAT tyAcgGTAAA AAGAATAAaG CAAGATTtAT TTCATTAGTA CTAATTTGTG 60
 CAATGTTTGC AATTGTGTTGG GTTGCATATA TTCAATGGGA GTCTACAATC GCTTCATTTA 120
 20 CACAATCTAT TAATATTTCa ATGGCACAAT ATAGTGTTTT ATGGACAATT AACGGAATAA 180
 TGATTTTAGT AGCACAACCA TTAATTAAAC CGATTCTCTA TCTGTTAAAA GGAAACTTAA 240
 AGAAGCAAAT GTTTGTCGGC ATCATCATTT TTATGTTGTC GTTCTTTGTC ACGAGTTTGT 300
 25 CCGAAACTT TACAATATTT GTTGTGGTA TGATTATTTT AACTTTTGA GAAATGTTTG 360
 TATGGCCAGC AGTTCCAATC ATAGCCAATC AGTTAGCGCC AGATGGTAAG CAAGGACAGT 420
 ACCAAGGTTT TGTGAATTCA GCTGCTACAG TAGGAAAAGC ATTTGGTCCA TTTCTTGGTG 480
 30 GTGTATTAGT TGATGCGTTT AATATGCGCA TGATGTTTAT CGGTATGATG CTACTACTTG 540
 TATTTGCATT AATATTATTA ATGGTTTTCA AGGAGAATAA TACGCAACCT AAAAAATAG 600
 ATGCATAATG AGTAAATAGA ATTAACGTTA TAGACTTGAA ATAAATGTCG TTATAACATA 660
 35 ATATTAATTT GTATAATTTA ATTTGTTTTG GAGCTTTTCT ACAGAAAGCT AGTGATGCTG 720
 AGAGCTAGTG TTAAGGACTA AATGTAAATC GTATTAATTT TAAATTGAAT GAATGACATC 780
 TCTTACTATT AAAATGAGTG CACAATTTTT GTGAAATAGG GTGGTAACGC GGCAAATGTC 840
 40 GTCCCTATGT AAATAGAATA GTTAGAGGTG TCTTTTTTAT TGAATAGGAG GAAATGTGTT 900
 GAATTACAAC CACAATCAAA TTGAAAAGAA ATGGcAAGAC TATTGGGACG AAAATAAAAC 960
 45 ATTTAAACA AATGATAACT TAGGTCAAAA GAAATTTTAT GCTTTAGACA TGTTTCATA 1020
 TCCATCAGGT GCTGGTTTAC ATGTTGGACA TCCTGAGGGc TATACAGCAA CAGATATCAT 1080
 TTCAAGATAT AAAAGAATGC AAGGATATAA TGTATTACAT CCGATGGGGT GGGATGCATT 1140
 50 CGGATTACCA GCAGAGCAAT ATGCTTTAGA CACTGGCAAC GACCCACGTG AATTTACAAA 1200
 GAAAAATATC CAAACTTTTA AACGACAAAT TAAAGAATTA GGGTTCAGTT ATGATTGGGA 1260

55

	GTTATATAAC AAAGGTTTAG CATACGTTGA TGAAGTTGCA GTTAACTGGT GTCCAGCATT	1380
	AGGCACTGTT TTATCTAACG AAGAAGTGAT TGATGGTGTC TCTGAACGTG GTGGACATCC	1440
5	AGTTTATCGT AAGCCGATGA AACAAATGGGT ACTTAAAATC ACAGAATATG CAGATCAATT	1500
	ATTAGCAGAT TTAGATGATT TAGATTGGCC TGAGTCTTTA AAAGATATGC AGCGCAATTG	1560
	GATTGGACGT TCTGAAGGGG CCAAAGTTTC ATTTGATGTA GATAATACGG AAGGAAAAGT	1620
10	AGAAGTATTT ACGACTAGAC CAGATACAAT CTATGGTGCA TCATTCTTAG TCTTAAGTCC	1680
	TGAACATGCA TTAGTTAATT CAATTACAAC AGATGAATAT AAAGAAAAAG TAAAAGCTTA	1740
15	TCAAAACAGAA GCTTCTAAAA AGTCAGATTT AGAACGTACA GATTTAGCAA AAGATAAATC	1800
	AGGTGTATTT ACTGGTGCAT ATGCAACTAA TCCTTTATCT GGTGAAAAAG TACAAATTTG	1860
	GATTGCTGAT TATGTATTAT CAACATATGG TACTGGAGCA ATTATGGCAG TACCAGCGCA	1920
20	TGATGACAGA GATTATGAAT TTGCTAAAAA GTTTGATTTG CCAATCATTG AAGTCATCGA	1980
	AGGTGGAAT GTTGAAGAAG CAGCATACAC TGGTGAAGGT AAACATATTA ATTCTGGTGA	2040
	ACTTGATGGT TTAGAAAATG AAGCGGCAAT TACTAAAGCT ATTCAATTAT TAGAGCAAAA	2100
25	AGGTGCTGGC GAAAAGAAAG TTAATTACAA ATTAAGAGAT TGGTTATTCA GTCGTCAGCG	2160
	TTATTGGGGC GAACCAATTC CTGTCATTCA TTGGGAAGAT GGAACAATGA CAACTGTTCC	2220
	TGAAGAAGAG CTACCATTGT TGTTACCTGA AACAGATGAA ATCAAGCCAT CAGGGACTGG	2280
30	TGAGTCTCCA CTAGCTAATA TTGATTCAAT TGTAATGTT GTAGATGAAA AAACAGGTAT	2340
	GAAAGGACGT CGTGAAACAA ATACAATGCC ACAATGGGCA GGTAGTTGTT GGTATTATTT	2400
	ACGTTACATC GATCCTAAAA ATGAAAATAT GTTAGCAGAT CCTGAAAAAT TAAAACATTG	2460
35	GTTACCTGTT GATTTATATA TCGGTGGAGT AGAACATGCG GTTCTTCACT TATTATATGC	2520
	AAGATTTTGG CATAAAGTCC TTTATGATTT GGCTATCGTA CCTACTAAAG AACCTTTCCA	2580
40	AAAATTATTT AACCAAGGTA TGATTTTAGG AGAAGGTAAT GAGAAGATGA GTAAATCTAA	2640
	AGGAAATGTA ATCAATCTG ATGATATAGT ACGTCTCAT GGTGCAGATA CTTTGCGTCT	2700
	TTACGAAATG TTTATGGGAC CTTTAGATGC TGCAATTGCA TGGAGTGAAA AAGGATTAGA	2760
45	TGGGTCTCGT CGATTCTTAG ATCGCGTATG GCGTTTAATG GTAAATGAAG ATGGGACATT	2820
	GAGTTCAAAA ATTGTAAC TAATAATAA ATCTTTAGAT AAAGTTTATA ACCAACTGT	2880
	TAAAAAGGTA ACAGAAGACT TTGAAACATT AGGATTTAAT ACTGCTATTA GTCAATTAAT	2940
50	GGTATTTATT AATGAGTGTT ATAAAGTTGA TGAAGTTTAT AAACCTTACA TTGAAGGCTT	3000
	CGTTAAAATG TTAGCACCTA TTGCACCACA TATCGGTGAA GAATTATGGT CAAAATTAGG	3060

55

	TGATGAAGTA GAAATCGTTG TTCAAGTGAA TGGTAAATTG AGAGCTAAAA TTAAAATTGC	3180
	TAAAGATACA TCAAAAGAAG AAATGCAAGA AATTGCCTTA TCTAATGACA ATGTTAAAGC	3240
5	GAGTATTGAA GGTAAAGACA TCATGAAAGT CATCGCTGTT CCTCAAAAAT TAGTCAATAT	3300
	TGTAGCTAAA TAATGTTTTA AGGAGGACTT TGAAATGAAG TCAATTACTA CAGATGAATT	3360
	AAAAAATAAA CTTTTAGAAT CTAAACCAGT TCAAATTGTT GATGTTCGTA CTGATGAAGA	3420
10	AACAGCAATG GGATATATTC CTAATGCAAA GTTAATTCCA ATGGATACCA TTCCGGATAA	3480
	TTTAAATTCA TTTAATAAAA ATGAAATATA TTATATTGTA TGTGCTGGTG GAGTTCGAAG	3540
	CGCTAAAGTT GTAGAATATT TAGAGGCAAA TGGCATTGAT GCCGTAAATG TCGAAGGCGG	3600
15	CATGCACGCA TGGGGCGATG AAGGTTTGGG AATAAAAAGT ATTTAAAGTA GTGACATAAT	3660
	TTAAAATAAT ATTACATTTG TAATGACACC AAGTAACGTT TCGGTTGCTT GGTGTTTTTT	3720
20	GGTATGAATT ACTTCTGTT ACAAACAAT CTAAAGCGTT CTTGTTATGT TTTATTAAGA	3780
	TTTTAATTAC AAAACGGAAA CTAAATTGTA ATAAAATAAA ACTTTATTTT ATAAAATGAT	3840
	GATGATAAAA TTGAGTGAAC TTAAAATATT GTACAAAATA ATATAGCTAT AAATATAATA	3900
25	TAGCTATAAA TATAATATGA GGGAGCGTAT ATTTTGTAGCA TAATTCTTAA CAACACAGCA	3960
	GAGAACAGAC AACCAGGAGG AAAATGAAAT GAATTTGTGA AAGAAAAATA AATATAGTAT	4020
	TAGGAAGTAT AAAGTAGGCA TATTCTCTAC TTAAATCGGA ACAGTTTAT TACTTTCAAA	4080
30	CCCAAATGGT GCACAAGCCT TAACTACGGA TAATAATGTA CAAAGCGATA CTAATCAAGC	4140
	AACACCTGTA AATTCACAAG ATAAAGATGT TGCTAATAAT AGAGGTTTAG CAAATAGTGC	4200
	GCAGAATACA CCTAATCAAT CTGCAACAAC CAATCAAGCA ACGAATCAAG CATTGGTTAA	4260
35	TCATAATAAT GGTAATATAG TAAATCAAGC TACGCCAACA TCAGTGCAAT CAAGTACGCC	4320
	TTCAACACAA AACAATAATC ATACAGATGG CAATACAACA GCAACTGAGA CAGTGTCAAA	4380
	CGCTAATAAT AATGATGTAG TGTGCAATAA TACCGCATT AATGTACCAA CTAAAACAAA	4440
40	TGAAAATGGT TCAGGAGGAC ATCTAACTTT AAAGGAAATT CAAGAAGATG TTCGTCATTC	4500
	TTCAAATAAA CCAGAGCTAG TTGCAATTGC TGAACCAGCA TCTAATAGAC CGAAAAAGAG	4560
45	AAGTAGACGT GCGGCACCGG CAGATCCTAA TGCAACTCCA GCAGATCCAG CGGCTGCAGC	4620
	GGTAGGAAAC GGTGGTGCAC CAGTTGCAAT TACAGCGCCA TATACGCCAA CAACTGATCC	4680
	TAATGCCAAT AATGCAGGAC AAAATGCACC TAACGAAGTG CTGTCATTTG ATGACAATGG	4740
50	TATTAGACCA AGTACCAACC GTTCTGTGCC AACAGTAAAC GTTGTTAATA ACTTGCCGGG	4800
	CTTCACACTA ATCAATGGTG GCAAAGTAGG GGTGTTTAGT CATGCAATGG TAAGAACGAG	4860

55

	TCGTATACAT	GGAAGTATA	CGAATGACCA	TGGCGATTTT	AATGGTATCG	AGAAAGCATT	4980
	AACAGTAAAT	CCGAATTCTG	AATTAATCTT	TGAATTTAAT	ACAATGACTA	CTAAAAACGG	5040
5	TCAAGGCGCA	ACAAATGTTA	TTATCAAAAA	TGCTGATACT	AATGATACGA	TTGCTGAAAA	5100
	GACTGTTGAA	GGCGGTCCAA	CTTTGCGTTT	ATTTAAAGTA	CCTGATAATG	TGAGAAATCT	5160
	CAAAATTCAA	TTTGTACCTA	AAAATGACGC	AATAACAGAT	GCGCGTGGCA	TTTATCAACT	5220
10	AAAAGATGGT	TACAAATACT	ATAGCTTTGT	TGACTCTATC	GGACTTCATT	CTGGGTCACA	5280
	TGTTTTTGT	GAAAGACGAA	CAATGGATCC	AACAGCAACA	AATAATAAAG	AGTTTACTGT	5340
	AACAACATCA	TTAAAGAATA	ATGGTAATTC	TGGTGCTTCT	CTAGATACAA	ATGACTTTGT	5400
15	ATATCAAGTT	CAATTACCTG	AAGGTGTTGA	ATATGTGAAC	AATTCATTGA	CTAAAGATTT	5460
	TCCAAGTAAC	AATTCAGGCG	TTGATGTTAA	TGATATGAAT	GTTACATATG	ATGCAGCAAA	5520
20	TCGTGTGATA	ACAATTAAAA	GTAAGTGGAGG	AGGTACAGCA	AACTCTCCGG	CACGACTTAT	5580
	GCCTGATAAA	ATACTCGATT	TAAGATATAA	ATTACGTGTA	AATAATGTGC	CGACACCAAG	5640
	AACAGTAACA	TTTAACGAGA	CATTAACGTA	TAAAACATAT	ACACAAGATT	TCATTAATTC	5700
25	AGCTGCAGAA	AGTCATACTG	TAAGTACAAA	TCCATATACT	ATCGATATCA	TCATGAATAA	5760
	AGATGCATTA	CAAGCCGAAG	TTGACAGACG	TATTCAACAA	GCTGATTATA	CATTTGCGTC	5820
	ATTAGATATC	TTTAATGGTC	TGAAACGACG	CGCACAAACG	ATTTTAGATG	AAAATCGTAA	5880
30	CAATGTACCA	TTAAATAAAA	GAGTTTCTCA	AGCATATATT	GATTCATTAA	CTAATCAAAT	5940
	GCAACATACG	TTAATTCGAA	GTGTTGATGC	TGAAAATGCA	GTTAATAAAA	AAGTTGACCA	6000
	AATGGAAGAT	TTAGTTAATC	AAAATGATGA	ATTGACAGAT	GAAGAAAAAC	AAGCAGCAAT	6060
35	ACAAGTTATC	GAGGAACATA	AAAATGAAAT	AATTGGTAAT	ATTGGTGACC	AAACGACTGA	6120
	TGATGGCGTT	ACTAGAATCA	AAGATCAAGG	TATACAGACC	TTAAGTGGGG	ATACTGCAAC	6180
40	ACCGGTTGTT	AAACCAAATG	CTAAAAAAGC	AATACGTGAT	AAAGCAACGA	AACAAAGGGA	6240
	AATTATCAAT	GCAACACCAG	ATGCTACTGA	AGACGAGATT	CAAGATGCAC	TAAATCAATT	6300
	AGCTACGGAT	GAAACAGATG	CTATTGATAA	TGTTACGAAT	GCTACTACAA	ATGCTGACGT	6360
45	TGAAACAGCT	AAAAATAATG	GCATCAATAC	TATTGGAGCA	GTTGTTCTCT	AAGTAACTCA	6420
	TAAAAAAGCT	GCAAGAGATG	CAATTAACCA	AGCAACAGCA	ACGAAAAGAC	AACAAATAAA	6480
	TAGTAATAGA	GAAGCAACTC	AGGAAGAGAA	AAATGCAGCA	TTGAACGAAT	TAATCAAGC	6540
50	AACCAACCAT	GCTTTAGAAC	AAATCAATCA	AGCAACAACA	AATGCTAATG	TTGATAACGC	6600
	CAAAGGAGAT	GGTCTAAATG	CCATTAATCC	AATTGCTCCT	GTAAGTGTG	TTAAGCAAGC	6660

55

5 TGATGCGACT CAAGAAGAAA GACAAGCAGC AATTGACAAA GTGAATGCTG CTGTAAGTGC 6780
 AGCAAACACA AACATTTTAA ACGCTAATAC CAATGCTGAT GTTGAACAAG TAAAGACAAA 6840
 10 TGCGATTCAA GGAATACAAG CAATTACACC AGCTACAAAA GTAAAAACAG ATGCAAAAAA 6900
 TGCCATCGAT AAAAGTGC GG AAACGCAACA TAATACGATA TTTAATAATA ATGATGCGAC 6960
 GCTCGAAGAA CAACAAGCAG CACAACAATT ACTTGATCAA GCTGTAGCCA CAGCGAAGCA 7020
 15 AAATATTAAT GCAGCAGATA CGAATCAAGA AGTTGCACAA GCAAAAGATC AGGGCACACA 7080
 AAATATAGTA GTGATTCAAC CGGCAACACA AGTTAAAACG GATACTCGCA ATGTTGTAAA 7140
 TGATAAAGCG CGAGAGGCGA TAACAAATAT CAATGCTACA ACTGGCGCGA CTCGAGAAGA 7200
 20 GAAACAAGAA GCGATAAATC GTGTCAATAC ACTTAAAAAT AGAGCATTAA CTGATATTGG 7260
 TGTGACGTCT ACTACTGCGA TGGTCAATAG TATTAGAGAC GATGCAGTCA ATCAAATCGG 7320
 CGCAGTTCAA CCGCATGTAA CGAAGAAACA AACTGCTACA GGTGTATTAA ATGATTTAGC 7380
 AACTGCTAAA AAGCAAGAAA TTAATCAAAA CACAAATGCA ACAACTGAAG AAAAGCAAGT 7440
 GGCTTTAAAT CAAGTGGATC AAGAGTTAGC AACGGCAATT AATmATATAA ATCAAGCTGA 7500
 25 TACAAATGCG GAAGTAGATC AAGCGCAACA ATTAGGTACA AAAGCAATTA ATGCGATTCA 7560
 GCCAAATATT GTTAAAAAAC CTGCAGCATT AGCACAAATC AATCAGCATT ATAATGCTAA 7620
 ATTAGCTGAA ATCAATGCTA CACCAGATGC AACGAATGAT GAGAAAAATG CTGCGATCAA 7680
 30 TACTTTAAAT CAAGACAGAC AACAGCTAT TGAAAGTATT AAACAAGCTA ACACAAATGC 7740
 AGAAGTAGAC CAAGCTGCGA CAGTAGCAGA GAATAATATC GATGCTGTTC AAGTTGATGT 7800
 AGTAAAAAAA CAAGCAGCGC GAGATAAAAT CACTGCTGAA GTGGcGAacG TATTGaAGCG 7860
 35 GTTAAACAAA CACCTAATGC AACTGACGAA GAAAAGCAGG CTGCTGTAA TCAAATCCAA 7920
 TCAACTTTAA AGATTCAAGC AATTTAATCC AAATTTAATC CAAAACCCAA ACAAATGGAT 7980
 TCAGGGTAGG ACACCACTTA CAAATCCAA 8009

(2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10953 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

ACCCACCCCh TGGGGATAnT TTACCTGGTG GGGCCTTCGA TTGCCTTTAG GTGAAACCaG 60

	AGATGAATGC TAACCATATT CATTCTGCTA AAGATGGTCG TGTTACTGCG ACAGCTGAAA	180
	TTATTCATCG AGGTAAGTCG ACACATGTAT GGGATATAAA AATTAAGAAT GACAAAGAAC	240
5	AATTAATTAC AGTTATGCGT GGTACAGTTG CTATTAAACC TTTAAAATAA AAGAACTGCT	300
	AGCTGAAATG TTATGAGATA TTCATAACTA CGGCTAGCAG TTTTTTTATG CGCTATATTG	360
	TTGTAGTTTT AGAAATGCTT GTTCAATGCG TTCGGCAGCT TTACGGCCAC CCATAACATT	420
10	TCTACCAAAT GGTCCTAATT CTAAGTCTGC AAAGCATCCT GCGACAAATA GATTTGGTAT	480
	CCATTCTAAT TTTTCGGAAA TAACAGGGTA ATTACATTCG TTGATAGGTG CATCATAATT	540
	TTGTATTAAT TGCTTAATAA GTGGTTGTGA CATAAAATCT TGTTCAAAAC CAGTTGCAAC	600
15	CATAATCTGT TGATATGGAA CAGAATCATT TTCAGTGTTA ATTACACCAC CACTAATTG	660
	AGTGATAGGT GTTTTATGCa CATTATACG ACCATTTTTTA ATATGTTTTT TAAGGCGTAA	720
20	GTACAGTTCG TGAGGCATTG ATCCTTTATG ACGTTCGCGT TGTACAATGG CATTTCTTTC	780
	AGGCATGCTT TTAGTACTTA AAAATGAAGA CATATTTTTTC GGACCTAACC AACCAGGATC	840
	AGCATCAAAG TCATGTATTT CAATATCTTT ATTTAGCCAT AAATGAATCT TTTTATCGTT	900
25	ATCATGATTT AACAAITTA GTGCAAGATG TGCAGCAGTa ATGCCGCTAC CAACGATATG	960
	ATCGGTCTTA TCATATACTA CTTGATCAAG TTCTTTCTCG AAGATATGAT TTACATTCTG	1020
	TTTGTCTTTT AAAATGTCAG GCATAAACGG AATATTTGTA CTGCCTATTG CAATAACGAC	1080
30	GCAATCTGTA GTGATAATTT GTCCATCTTC TAACTTGATA TGCCATTGT CTCTCTGTTT	1140
	ATCTAAAGTT TGAATAAAC CTTGAACCAA GCAATCCTCT AATTGATATT GTTTAGAAGC	1200
	ATGTGCAATA TGATCCATAA ACATTGTCAA TTCAGGTCGT TGATAAGGAC CATAAAAAGC	1260
35	ATTTGTATAT TGGTGTCTT TAGCGAATTG TTTTAGATGG AACGGTTGTG GATGTACGTG	1320
	ATGTACAATC GGTGATCTTA AATAAGGCAT TTCTATTGCA TTTGTATATG AGTTAAACCT	1380
	TTGGCAAAAA GTTTCGTGTG GGTCAATGAT TGTTAATCGG TCTGTTGTTA ATCCGCTTGA	1440
40	TAATAGTTTT TGTGCGATTG CAGTTCCCTG TATGCCACCG CCGATAATTG TCCAATGCAT	1500
	AATAAAACCT CTCTCTTTTT AAAACGTAAT AGTTACGATT TATAATTATT ATTATCATAA	1560
45	TACATAACGA CATGAAAGGC AATTAAATTA AAGAGATATA TGTAGATAGG GCGAATCTGT	1620
	AGTCAAAGAA AAAATCATTG AAAAAGAGGT AACAATGTCA AAAGAWAACA GCAGTAAAT	1680
	CATTCTAAT TTGGAATCAT CTTACTGCTG TTTGTTGTTG ATTTATATTC ATGATTTTGT	1740
50	TATATAATCT ACAATTTTGT GTCTTTTAAG TCTTCCGAAA TTTCATCGAC TTTAGTCTTT	1800
	TTAGTATAAG GCGTTTTAAT ATTATATGCT GCTTTCATAA TCATATGACT TGAAAGAGGA	1860

55

	GCAATAAAAT ATAAAAACGT ACCAAATAGT AATGACATTG CACCTAATGT TGATGCTTTT	1980
	CCGGCAGCAT GTGCACGTGA ATATACATCT TCAAGTCTCA ATAATCCTAT AGCTGCTAGG	2040
5	GCGCTAATTA AAGCACCGAT GATAACAAAG ATAAGTGCAA GACTAATCAG TATGATTTTG	2100
	ATCATGTTCA ATCACCTTAC CTTTGTCCAT AAATTTAGAG AATACTGCAG TACCTAAAAA	2160
	AGCTAATATA CCAATCATCA TAATAACGAC AATCATGTAT TTAATATTTA ATAAAATACT	2220
10	GAATAATGCT ATAACTGCCA TTAATTGAAG ACCAATCGCA TCTAATGCGA CAACACGATC	2280
	GGCAAGTGAT GGGCCTAGCA CAACGCGAAT GAGCATAGCT AACATAGAAA TGACAACTAT	2340
	GATTAATGCA ATAACGATAA TAACATTATG ATTCATTATA TTTCGCCCAC CTCTCTTACA	2400
15	ATTTTCTCTA ATGATGTTTT AATACTTTCT ACTTCTTGCT CTTTAGTTGA AAAATCTATG	2460
	GCATGAATAT AAATTTTTGT ACGATCGTCA CTTACACCAA GCACTACAGT ACCAGGTGTT	2520
	AATGTAATTA AATTAGACAG CAAGACAATT TGCCAATCTT TTTTAAATC TGTGTGATAA	2580
20	ACAAAGAATC CTGGTTCATT TTTAATCGAA GGTTTAATAA TAATTTTCAA AACATCAAAA	2640
	TTAGCTTTAA TCAGTTTCGAT TAAGAAAATA ATAATAATT TAATAATACG ATATAGCGTG	2700
25	ATGACATAAA ATCTACCTGG TAACACTCTG TGTAAGAGGT AAACAAGAAC TAGGCCAAAG	2760
	ATGAAACCTA ACACAAAGTT ATTTGTTGTG TAACATTTTG TCACAAACAA CCAAAACACT	2820
	GCGATAATAA AGTTTAATAC TAATTGTACA GCCATGTTAT TTACCTCCTA ATACAGCTTT	2880
30	AACGTAGGTT GATGGATTGT AGAATGTTTC TGCACCAGCT TTTACCATTG GATATAAGTA	2940
	ATCTGCTGAC AATCCATATA AAACAGTTAT CACAACTGCA ACGATTGCAA TCGTAGTTAA	3000
	ATATTTGACG TCGACTTTGT TATTAAGATC ATATCCTTTT GGTTGACCGA AAAAGCCTTG	3060
35	TAGGAATATG CGAATGACAG AATATAATAC GACTAAACTT GATAATAAGA CGATGACACC	3120
	ACTTAAATAA AATCCTCTTT CAAATGTTGA TTGGACAATA AAAAATTTTC CATAAAAGCC	3180
	ACTGAGTGGG GGAATGCCAG CTAAACTTAA TGCTGCGATA AAGAATGACC AACCAAGTAC	3240
40	AGGATATCGT TTAATTAAGC CACCAAATTG TCTTAAATCA GCAGTGCCTG TAATTTTAAT	3300
	CATAATTCCG ATAAGCAAGA ATAATGCAAG TTTTACTAAC ATGTCGTGCA ATGTATAGTA	3360
45	AATAGCCCCA ATCATACCTG ACTCTGTCAT CATTGCAACG CCGACTAAGA TCACACCTAC	3420
	AGCAATCATG ACATTGTATA GGATGATTTT TTTAATGTTG GCATATGCAA CAGCACCGAC	3480
	ACAACCAAAG ATGATCGTTA ATAGTGCTAA GAATAAAATG ACATAATGTG AAAAGCTTAC	3540
50	ATTATCACTA AAGAATAGGC TCAATGTTCT AGCGATTGCA TAAACACCAA CTTTTGTTAA	3600
	CAAAGCACCA AAGAATGCAA TGATTGGAAT TGGTGGCAT AGTATGCACT AGGTAACCAA	3660

55

	ATATTGACTA AGCCACTGTC ATGCGCTGAA AGGTTAGCTA ATTTATTGCT TATATCTGCT	3780
	AGATTCAATG TTCCTACTAC TGAATATAAA ATCGCTACAC CCATTACGAA GAAGGATGAC	3840
5	GATACAACGT TAACAAGAAC ATATTTTATT GTTTCTTGTA GTTGAATTTT TGTAGAACCA	3900
	ATTACTAATA AGAAATAAGA TGACATTAAA AATACTTCGA AAAATACGAA TAGGTTGAAA	3960
	ATGTCACCAG TTGTGAATGC ACCAATGATA CCTATTAACA TAAATAGTAC TGAAAAATAA	4020
10	TAATAATATC TTTCACGTTT AATACCAATT GTTTGGTATG AATATAAAAT CACAATAGCT	4080
	GTAATAATAA TACTAGTAAT TATTAGTAGG GCACTGAATA TGTCTAATAC AAAGACAATA	4140
	CTGTATGGTG CTTTCCATGA ACCTAGCTCT ACGCGTATTG GTCCATGTTT AACCAACATTT	4200
15	GCTAAATTGA TAATTGCCGC GACCAAGGTT AATAATGTAC CGCCTAGTGC GACATAACGC	4260
	TTTATAATAG GACGCTTTCC AATAAAGACA AGTAATATGG CTGTAATTAC TGGAATAACT	4320
	AGCGTTAACA CAAGCATATT ACTTTCAATC ATCTTCTGGA ACTCCTTTCA TACTCTCAAC	4380
20	GTTATCTGTG CCTAATTCTT TATATGTTCT AAATGCTAAT ACTAAGAAAA AGGCTGTTGT	4440
	CGCAAGCGCA TAACGATTGC TGTTAAAATA AGTGCTTGCG GGaTAGGaTC AACATAGCTT	4500
25	TTTACGTTTCG CTTTATAAAT TGGAACAGTA CCATGTTTAA GTCCGCCCAT AGTTATTAAA	4560
	AATAAATTTG CTGCATGTGT TAATAGTGTA GTTCCCATAA CAATTCGTAT CAGACTTTTA	4620
	GACAAAACGA GATAGACACT AATTGCTGTG AGAATACCAC TAACAAAAAT CATAATAATT	4680
30	TCCACTATTG GTTCTCTCCA ATCGAAATAA TAATGTGCAT GACAGTACCA ACTACTGCAC	4740
	ATAAAACACC GAAATCAAAG AATACTGCTG TTGTCATATG AACAGGTTCT AATATAAATA	4800
	ACGGTATATC AAATGTGACA TCGGTAAAGA AATTTTGGCC TAAAAACCA CTTGCGATAG	4860
35	GCGTCGCAAT ACAAAAAACT AATCCGATAC CTATCAAGAT TTTAAATCT AATGGGAAAA	4920
	TTTTACGCAT TGTTTCTATA TCAAATGCAA TCGTAATGAT AACAAAGTGAA CTTGCGAATA	4980
	ATAATCCGCC GACGAAACCG CCACCAGGTG TATAATGTCC TGCTAAGAAA AGTGAAAAAC	5040
40	CAAAGACCAT TACCATGAAA AAGATAATAA CTGCAGCAAA TGCAAAATT AGATCATTTT	5100
	GTTGTCTATT CATGATTTTT CACCTCGTTA CCTTGCCTTT GACGCTTTTT ACGTAATTTA	5160
45	ATCATTGTAT ATACAGCTAA TCCTGCGATA CCAAGCACAG ATGACTCGAA TAAAGTATCC	5220
	ATACCACGGA AATCAACAAG TATGACGTTT ACCATGTTTT TACCGTGAGC tAAATCATAA	5280
	ACGTGCTCTT GATAAACTT AGATATCGAT TCAAAATGTC TATTTCCGTA TGCAATTAAA	5340
50	CCGATAATAA TGACGGACAA ACCAACACCA CCAGCAATTA AAGCATTAGT AAGCTGGAAT	5400
	GAGCGCTTTT CATTATAACG ATTTAAATTT GGTAAGTGGT AGAAGCATAA TAAGAACAAT	5460

55

	ATAACAATA CAGACACAGC ATATCCAAC TCACTTAACA TAATGATGCT AAATAATCTT	5580
	GATTTAGCGA AAAGAATTAA AAAGGCAGCA CTTAATAATA AAATTACGAT ACAAACCTCG	5640
5	AAAATTCTAA TCGGACTAAC GTCTTTAAAA TTAATGTTGA AAGGTACTGA GAATATAGTG	5700
	ACAAATGTTA ATAAAAATTAA TGCACCAAAA ATGATAACTA AATTATTACG TGAATAATCG	5760
	GTAACATAGC TATTCGTCAT CTTTTCAGAG TAGTTTGGAA TAACATTTGC ACTTCTGTTG	5820
10	TACCAATAAT TGAATGTTAG TTTACCAGGT TGTCGTTGCA ACAATTTTAC CCAATAACTA	5880
	AATGTCACAA TTAGTAAGAT ACCTAAAAATA TAAATCACTA ATGTTGATAA AAAGGCAGGC	5940
15	GTTAATCCAT GGAACATATG GAATTCAACA TCATCAATTA CCGTATGATT AATCGAAGag	6000
	TnAGCTGGTT CAATAATCGA ATTAGTTAAA ATGCCAGGGA ATAAACCAAA TACAATTACT	6060
	AATGTAGCTA AAATAGCTGG TGATAAAAGC ATTAATATTG ATACTTCGTG TGCTTTTTTA	6120
20	GGTAATTGTT CAGGTTTATA TTGTCCGAAA AATATATGCA TTATAAATTT AATTGAATAT	6180
	ACAAATGTGA AGACACTGCC CACTATACCA ATGATTGGGA ATAGGTAGCC TAATGTATCA	6240
	ACACTGAATA AATTGCTTG GCTTGCTGTA AATGTTGTTT CTAAAAATGA TTCTTTTGAT	6300
25	AAGAAACCAT TGAACGGTGG TACACCAGCg CATACTTAAT GCTGTAATAA CAGTGATTGT	6360
	AAATGAAATA GGCATAATTG TTAGTAAGCC ACCTAATTTT TTAACATCAC GTGTACCAGT	6420
	AGAATGATCC ACTGCACCTG TAATCATAAA TAGGGCACCT TTAAATGTTG CATGGTTGAT	6480
30	TAAATGGAAT ATTGCAGCCG TAAATGCAGC AGCATATATT TTGCTATCAT CGCCTTGATA	6540
	GTGATAACTA ATGGCACCGA TTCCAAGCAT CGCCATAATC ATACCTAATT GGGATACTGT	6600
	TGAAAATGCC AGTATACCTT TCAAGTCTTG TTGTTTTGTT GCGTTTAGCG AAgCCCAGAA	6660
35	TAATGTAATT AAACCAACGA GTGTGACAGT CCATACCCAA CCTTGCGATG CTGCGAAGAT	6720
	TGGTgTCATT CGAGCGATTA AATATAACCC TGCTTTAACC ATTGTTGCTG AATGAAGATA	6780
40	AGCACTGACT GGTGTAGGTG CTTCCATTGC ATCTGGTAGC CAAATATAAA ATGGAACTG	6840
	AGCAGATTTT GTAAAAGCAC CAATCATGAT TAAATCATC GCAAAAATGA AGAATGGGCT	6900
	ATTTTGAATT TCAGAAGCAT GTTGAATCAT GTACTGAATG CTAAATGATT GTGTTGGTAT	6960
45	AGCGAGTAAG ATGATACCAC CTAATAATGA TAGACCACCA AATACTGTGA TTATGAGCGA	7020
	TTTTTGAGCA CCATATATAG ATGCTTGTCG TTCGCGCCAG AATGAAATAA GTAAAAAACT	7080
	AGAAAATGAC GTTAGCTCCC AGAATAAATA TAGAATAATA ACATTATCTG AAAGTACGAC	7140
50	ACCTAACATT GCACCCATAA ATAGTAATAA ATAACAATAA AAATTCCTTA GTTGTCTGA	7200
	CTTACTTAAG TAGCCGATTG AATATAATAC TACTAACTG CCGATTCTG AAATAAGCAA	7260

55

	CCAATTTAAG GTTTTCATTA CAGTATTACC TGACATCGTC GTTTTAATTA ATGTAAGCAT	7380
	ATAAATAAAT ATGACGATAG GGACAGGTAA TACGAACCAT CCTAAATGTA TACGTTTAAA	7440
5	AAATCTATAC AGGATAGGAA TAATGAGTGC GAATATTAAC GGTAATATCA CCGCAATATG	7500
	TAACAAACTC ACTATGTTGT CCTCCTTTAA AAAATATTTA TGTATTTCAT TATACATGAA	7560
	TGATATAGTT CTGAAAAACG TACACACTCC TTGTTGTGCT TTATTTTCAG AaGTATTTAA	7620
10	ATAAGAAGAA ACACGTCATT TTTTATTTAA AATTTTCTTT GTATTGAAGT GAATAATCTT	7680
	CTTTTAAGCG TGCTAAACTA GCTAAAGACA TTTCAGCATG TTTTGTTCG TGAGCTTTAA	7740
	GTTTAGTTTC TAAATCTGTA ATTGCTTGTT GAAGTGAATC TTCATAGCGC AATACATCAA	7800
15	CATTGAAGTC GCGTAATTGT GAACGTTTCG TATAGCGTTT TTCAAATGG CTTAATGCTT	7860
	TGCGGTCATG GAAAAATACA CCTTCAGTTT CAGTAGGGTT ATGTAAATCA CCTTGTTTCG	7920
20	GGTGTTCGAT AACTTGTTCA ACTTTAACAA GGACATCGTC TCCATTTTCT TCAACAATCG	7980
	TGACACCATA GCTACCTGTT TTGTGTGAAA ATCGATATAG CTTCATGCTA TTTTCCTCCC	8040
	TTAAAAGTAT GTTAATATAT ATGTATCATA ACATGAATGG AGAATATAAA TGGCTAACTA	8100
25	TCCACAGTTA AACAAAGAAG TACAACAAGG TGAAATCAAA GTGGTTATGC ACACAAATAA	8160
	AGGTGACATG ACATTCAAAT TATTTCCAAA TATTGCACCA AAAACAGTTG AAAATTTTGT	8220
	GACACATGCA AAAAATGGTT ATTATGATGG AATCACATTC CACCGTGTCA TTAATGACTT	8280
30	CATGATTCAA GGTGGCGATC CAACAGCTAC TGGTATGGGT GGCGAAAGTA TTTATGGCGG	8340
	TGCTTTTGAA GATGAATTTT CATTAAATGC ATTAACTTA TATGGCGCAT TATCAATGGC	8400
	TAACTCAGGA CCTAATACTA ATGGTTCACA ATTTTTCATT GTTCAAATGA AAGAAGTACC	8460
35	TCAAATATG TTAAGTCAAC TTGCAGATGG TGGCTGGCCT CAACCAATCG TTGATGCATA	8520
	TGGCGAAAAG GGTGGTACAC CATGGTTAGA TCAAAAACAT ACAGTATTCG GTCAAATCAT	8580
	TGATGGTGAA aCTACATTAG AAGATATTGC AAATACAAA GTGGGACCAC AAGATAAACC	8640
40	ACTTCATGAT GTTGTAATTG AATCTATTGA TGTGAAGAA TAATATCTAA ACATAATTAA	8700
	CTACCAACAT TTAAACTCG GATAAAGCTA ATTTATGAAT GGATTAGTAT ATATTCCAAC	8760
45	gAAAATAAAT AAATAATAT GATGAGCAAT CTCAATATAT TTATCaAGAA AGCACAGTTT	8820
	TTAAATAGAT GTGTATTTTA AAGATAATAG TTGAGGTTGC TTTTATGTT TTTACAGAGA	8880
	ATTGCTATTC AAATAGTAAA TAAATTGAAA ACAAAGTAGC TGGATATCAT ATTGATTAG	8940
50	ATAGGAATTT GTTGCTAATT TTATTTGTAA ATCCAAGTTT GTAGAATTCT TATTCATTTA	9000
	TAAATAATA TTCGTATGAT TTGATTTTTT AATTAGTCCA CCATTTTCGAT TTGTGCTATG	9060

55

AACATATCAA GGTGCGTGTA CTGGTATTCA ACCATACGGT GCGTTTGTG AGACCCCTAA 9180
 TCATACTGAA GGACTGATTC ATATATCAGA AATTATGGAT GACTACGTTC ATAATTTGAA 9240
 5 GAAATTTCTA TCAGAAGGCC AAATTGTTAA AGCTAAAATT TTGTCTATAG ATGATGAAGG 9300
 AAAGCTTAAT CTATCATTAA AGGATAATGA TTACTTCAAA AATTATGAGC GTAAGAAGGA 9360
 AAAACAATCA GTATTAGATG AAATCAGAGA AACAGAAAAA TATGGGTTTC AAACACTTAA 9420
 10 AGAACGCTTA CCAATCTGGA TAAAACAGTC AAAGCGAGCA ATTCGAAACG ACTAAAGGAA 9480
 CAGATAAATC GTACCGAAAA TCATACAAAG GGTCTGAAAT GAAAGTTTCT TAGACTATAA 9540
 AAGAGATTAG TATCTATTAA ATTTTATTAG ATACTAATCT CTTTTTGTCT ACGATAACGT 9600
 15 AATATGATTG ATTCTATTTA CACGTACAAA TGGTTTAAGG TGACATATCC ATTATCTTTG 9660
 TTAGATAGAA TCGTTGATTT GCaATATTGT ATGTGGATTT GTTTTTTTTA TTTATTTTAG 9720
 20 AAATGAGAAC TACAACTTAA AGTATTAAAC GAATTGCAAC TATATAAACA GATAATTGGA 9780
 GAATGAAAAA ATTACATGTT ATAGTCAACT CAATAATTTT AAGGAGGAAT TAAGTAATGA 9840
 AAAGTAAATA CGAACCATTG TTGATAAAG TAGAATTACC AAATGGAGTA GAGTTGAGAA 9900
 25 ATCGATTGT GTTAGCCCCT TTAACACATA TTTCTTCAAA TGATGATGGT ACTATTTTCA 9960
 ATGTAGAACT TCCTTATATT GAAAAGCGTT CACAAGATGT TGGTATTACA ATTAATGCTG 10020
 CGAGTAATGT GAGTGATGTC GGAAAAGCAT TTCCAGGACA GCCATCAATC GCGCATGACA 10080
 30 GTAATATTGA AGGACTAAAA CGATTAGCTA CAGCAATGAA GAAAAACGGT GCCAAAGCAC 10140
 TCGTACAAAT ACATCATGGC GGTGCACAAG CATTGCCTGA ATTAACACCT GATGGAGACG 10200
 TCGTAGCACC AAGTCCAATT TCTTTAAAAA GTTTTGGTCA GAAACAAGAA CATAGTGCTA 10260
 35 GAGAAATGAC GAATGAAGAG ATTGAACAAG CAATCAAGGA TTTTGGTGAA GCAACGCGAC 10320
 GTGCAATTGA AGCAGGGTTT GATGGTGTG AAATACATGG CGCGAATCAT TACTTAATTC 10380
 ATCAATTTGT ATCACCATAC TATAATAGAA GAAATGATGT ATGGGCAAAT CAATATAAAT 10440
 40 TCCCGGTGCG TGTGATTGAA GAAGTACTTA AAGCGAAAGA AGCGTATGGC AATAAAGACT 10500
 TTATAGTTGG ATACAGATTA TCTCCAGAGG AAGCGGAGTC TCCAGGAATC ACAATGGAAA 10560
 45 TTACAGAGGA ACTCGTTAAT AAAATTAGCC ATATGCCAAT CGACTATATT CATGTTTCAA 10620
 TGATGGATAC GCATGCAACG ACACGTGAAG GTAAATACGC TGGACAAGAA AGACTGCCTT 10680
 TAATTCACAA ATGGATAAAT GGTGCTATGC CACTTATCGG TATTGGTTCA ATTTTCACAG 10740
 50 CTGACGAAGC TTTAGATGCA GTTGAAAATG TTGGTGTGTA CTTAGTAGCC ATTGGTAGAG 10800
 AGCTACTACT GGATTATCAA TTTGTTGAAA AAATTAAAGA TGGACGGGAA GATGAAATTA 10860

55

AATTTAATGA AGGGTTTTAT CCATTACCAC GTA

10953

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8155 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTTGATAnAA AACTGAATnA ATTAAATGTA TCGATTCAAC CTAATGAAGT GAATTTACAA	60
GTAAAGTAG AGCCTTTTAG CAnAAAGGTT AAAGTAAATG TTAAACAGAA AGGTAGTTTA	120
GCAGATGATA AAGAGTTAAG TTCGATTGAT TTAGAAGATA AAGAAATTGA AATCTTCGGT	180
AGTCGAGATG ACTTACAAAA TATAAGCGAA GTTGATGCAG AAGTAGATTT AGATGGTATT	240
TCAGAATCAA CTGAAAAGAC TGTAAAAATC AATTTWCCAG AACATGTCAC TAAAGCACAA	300
CCAAGTGAAA CGmAGGCTTA TATAAATGTA AAATAAATAG CTAAATTAAA GGAGAGTAAA	360
CAATGGGAAA ATATTTTGGT ACAGACGGAg TAAGAGGTGT CGCAAACCAA GAACTAACAC	420
CTGAATTGGC ATTTAAATTA GGAAGATACG GTGGCTATGT TCTAGCaCAT AATAAAGGTG	480
AAAAACACCC ACGTGTACTT GTAGGTCGCG ATACTAGAGT TTCAGGTGAA ATGTTAGAAT	540
CAGCATTAAAT AGCTGGTTTG ATTTCAATTG GTGCAGAAGT GATGCGATTA GGTATTATTT	600
CAACACCAGG TGTTGCATAT TTAACACGCG ATATGGGTGC AGAGTTAGGT GTAATGATTT	660
CAGCCTCTCA TAATCCAGTT GCAGATAATG GTATTAAATT CTTTGATCA GATGGTTTTA	720
AACTATCAGA TGAACAAGAA AATGAAATTG AAGCATTATT GGATCAAGAA AACCCAGAAT	780
TACCAAGACC AGTTGGCAAT GATATTGTAC ATTATTCAGA TTACTTTGAA GGGGCACAAA	840
AATATTTGAG CTATTTAAAA TCAACAGTAG ATGTTAACTT TGAAGGTTTG AAAATTGCTT	900
TAGATGGTGC AAATGGTTCA ACATCATCAC TAGCGCCATT CTTATTTGGT GACTTAGAAG	960
CAGATACTGA AACAAATTGA TGTAGTCCTG ATGGATATAA TATCAATGAG AAATGTGGCT	1020
CTACACATCC TGAAAAATTA GCTGAAAAAG TAGTTGAAAC TGAAAGTGAT TTTGGGTTAG	1080
CATTTGACGG CGATGGAGAC AGAATCATAG CAGTAGATGA GAATGGTCAA ATCGTTGACG	1140
GTGACCAAAT TATGTTTATT ATTGGTCAAG AAATGCATAA AAATCAAGAA TTGAATAATG	1200
ACATGATTGT TTCTACTGTT ATGAGTAATT TAGGTTTTTA CAAAGCGCTT GAACAAGAAG	1260
GAATTAAATC TAATAAAACT AAAGTTGGCG ACAGATATGT AGTAGAAGAA ATGCGTCGCG	1320

	CTGGTGATGG TTTATTAAC T GGTATTCAAT TAGCTTCTGT AATAAAAAATG ACTGGTAAAT	1440
	CACTAAGTGA ATTAGCTGGA CAAATGAAAA AATATCCACA ATCATTAAAT AACGTACGCG	1500
5	TAACAGATAA ATATCGTGTT GAAGAAAATG TTGACGTTAA AGAAGTTATG ACTAAAGTAG	1560
	AAGTAGAAAT GAATGGAGAA GGTCTGAATTT TAGTAAGACC TTCTGGAACA aACCATTAGT	1620
	TCGTGTCATG GTTGAAGCAG CAACTGATGA AGATGCTGAA aGATTGTCAC AACAAATAGC	1680
10	TGATGTGGTT CAAGATAAAA TGGGATTAGA TAAATAAATA CTGTATTACA AATGAGCCGA	1740
	TGCGTATGcA nTcgtTTTT GTGTTTGTAG AAATAATTTA TAGTACAAAC GTAAAAATGAT	1800
	ATAAACAAAA TAAAAACAAA GTAATCAATA TGTAATATAA AATACACTGG TACTCAATAT	1860
15	ATAATGATGA TAAATTAAT TTTAATTAGA TAGAGTTGCT TTGTGTTTTT AACGCAGATG	1920
	CTACTACTTA TCTTAACAGT TGATTAAGTG AAATCATTTA ACAGCGAGAA TAATCAACCA	1980
20	GGAGGATGAC TTAATGAATT TATTCAGACA ACAAAAATTT AGTATCAGAA AATTTAATGT	2040
	CGGTATTTTT TCAGCTTTAA TTGCCACTGT TACTTTTATA TCTACTAACC CGACAACAGC	2100
	GTCTGCAGCA GAGCAAAATC AGCCTGCACA AAATCAACCA GCACAACCAG CTGATGCCAA	2160
25	TACACAGCCT AACGCAAATG CTGGTGCTCA AGCTAATCCT ACAGCACAGC CAGCTGCACC	2220
	TGCCAACCAA GGACAACCAG CAGTACAACC AGCAAACCAA GGTGGACAGG CTAATCCAGC	2280
	AGGAGGAGCA GCACAACCAA ATACACAACC AGCTGGACAA GGTGATCAAG CTGATCCGAA	2340
30	TAACGCTGCA CAAGCACAAC CTGGAAATCA AGCAACACCG GCAAACCAAG CAGGTCAAGG	2400
	AAATAACCAA GCAACACCTA ATAATAATGC AACACCGGCA AATCAAACAC AGCCAGCGAA	2460
	TGCTCCAGCA GCAGCGCAAC CAGCAGCACC TGTAGCAGCA AACGCACAAA CTCAAGATCC	2520
35	AAATGCTAGC AATACTGGTG AAGGCAGTAT TAATACGACA TTAACATTTG ATGATCCTGC	2580
	CATATCAACA GATGAGAATA GACAGGATCC AACTGTAACT GTTACAGATA AAGTAAATGG	2640
40	TTATTCAATTA ATTAACAACG GTAAGATTGG TTTCGTTAAC TCAGAATTAA GACGAAGCGA	2700
	TATGTTTGAT AAGAATAACC CTCAAAATA TCAAGCTAAA GGAAACGTGG CTGCATTAGG	2760
	TCGTGTGAAT GCAAATGATT CTACAGATCA TGGTAACTTT AACGGTATTT CAAAACTGT	2820
45	AAATGTAAAA CCAGATTCAG AATTAATTAT TAACTTTACT ACTATGCAAA CGAATAGTAA	2880
	GCAAGGTGCA ACAAATTTAG TTATTAAAGA TGCTAAGAAA AATACTGAAT TAGCAACTGT	2940
	AAATGTTGCT AAGACTGGTA CTGCACATTT ATTTAAAGTA CCAACTGATG CTGATCGTTT	3000
50	AGATTTACAA TTTATTCCTG ACAATACAGC AGTTGCTGAT GCTTCAAGAA TTACAACAAA	3060
	TAAAGATGGT TATAAATACT ATTCATTCAT TGATAATGTA GGTCTATTCT CAGGATCACA	3120

55

	TAATACTGAA ATCGGTAACA ATGGTAATTT TGGTGCTTCA TTAAAAGCAG ATCAATTTAA	3240
	ATATGAAGTA ACATTACCAC AAGGTGTAAC TTACGTTAAT AATTCATTAA CTACAACATT	3300
5	CCCTAATGGT AATGAAGACA GTACAGTATT GAAAAATATG ACTGTTAATT ATGATCAAAA	3360
	TGCAATAAAA GTTACATTTA CAAGCCAAGG TGTGACAACG GCACGTGGTA CACACACTAA	3420
	AGAAGTTTAA TTCCCAGATA AATCTTTAAA ATTATCATAT AAAGTTAATG TTGCGAATAT	3480
10	CGATACACCT AAAAAATATTG ATTTTAATGA AAAATTAACA TATCGTACTG CTTCAGATGT	3540
	TGTAATTAAT AATGCGCAAC CAGAAGTaCA CTAAGTGCAG ATCCATTTTC AGTAGCGGTT	3600
	GAAATGAACA AAGATGCGTT GCAACAACAA GTAAACTCAC AAGTTGATAA TAGTCATTAC	3660
15	ACAACAGCAT CAATTGCAGA ATACAATAAA CTTAAACAAC AAGCAGATAC TATTTTAAAT	3720
	GAAGATGCGA ATCATGTTAA AACTGCAAAT CGTGCATCTC AAGCGGATAT TGATGGTTTA	3780
20	GTAAGTAAAT TACAAGCTGC ATTAATTGAT AATCAAGCAG CAATTGCTGA ATTAGATACT	3840
	AAAGCTCAAG AAAAGGTTAC AGCAGCACAA CAAAGTAAAA AAGTTACGCA AGATGAAGTT	3900
	GCAGCACTTG TAACTAAAAT TAACAATGAT AAAAATAATG CAATCGCAGA AATTAATAAA	3960
25	CAAACTACAG CACAAGGTGT CACAAGTAA AAAGATAATG GTATCGCAGT GTTAGAACAA	4020
	GATGTGATTA CACCAACAGT TAAACCTCAA GCGAAACAAG ATATTATCCA AGCAGTTACA	4080
	ACTCGTAAAC AACAAATTAA AAAGTCAAAT GCATCATTAC AAGATGAAAA AGATGTAGCA	4140
30	AATGATAAAA TTGGTAAAT TGAAACAAAG GCAATTAAAG ATATTGATGC AGCAACAACA	4200
	AATGCACAAG TAGAAGCCAT TAAAACAAAA GCAATCAATG ATATTAATCA AACTACACCT	4260
	GCTACAACAG CTAAAGCAGC AGCTCTTGAA GAATTTGACG AAGTTGTTCA AGCACAAATT	4320
35	GATCAAGCAC CTTTAAATCC TGATACAACA AATGAAGAAG TAGCGGAAGC TATTGAACGT	4380
	ATTAATGCAG CTAAAGTTTC TGGTGTAAAA GCAATTGAAG CGACAACGAC TGCACAAGAT	4440
	TTAGAAAGAG TTA AAAACGA AGAAATCTCA AAAATTGAAA ATATTACTGA CTCTACGCAA	4500
40	ACAAAAATGG ATGCCTATAA TGAAGTTAAA CAAGCTGCAA CAGCTAGAAA AGCTCAAAT	4560
	GCTACAGTTT CAAATGCAAC AAATGAAGAA GTAGCAGAAG CTGATGCAGC AGTAGATGCA	4620
45	GCTCAAAAGC AAGGTTTACA TGACATCCAA GTTGTTAAAT CAAAACAGGA AGTTGCTGAT	4680
	ACAAAATCAA AAGTATTAGA TAAATCAAT GCAATTCAAA CACAAGCAAA AGTTAAACCT	4740
	GCAGCTGATA CGGAAGTAGA AAACGCATAT AATACACGTA AACAAGAAAT TCAAAATAGC	4800
50	AATGCTTCAA CTACAGAAGA AAAACAAGCT GCATATACAG AATTAGATAC TAAAAAGCAA	4860
	GAAGCAAGAA CAAATCTTGA TGCTGCAAAT ACAAACAGTG ATGTAACAAC AGCTAAAGAC	4920

55

	GCGGAAATCG CTCAAAAAGC AAGTGAACGT AAAACAGCAA TTGAAGCAAT GAATGATTCG	5040
	ACTACTGAAG AACACAAGC AGCGAAAGAC AAAGTGGATC AAGCAGTAGT TACTGCAAAC	5100
5	GCTGATATAG ATAATGCTGC AGCAAACAAT GATGTGGATA ATGCAAAAAC TACAAATGAA	5160
	GCTACAATCG CAGCCATTAC ACCTGATGCA AATGTTAAAC CAGCAGCAA ACAAGCAATT	5220
	GCAGATAAAG TACAAGCTCA AGAAACAGCA ATTGATGGAA ATAACGGCTC AACAACTGAA	5280
10	GAAAAAGCAG CTGCTAAACA ACAAGTTCAA ACTGAAAAAA CAACAGCTGA TGCCGCAATA	5340
	GATGCAGCAC ATACAAATGC GGAAGTTGAA GCGGCTAAAA AAGCAGCAAT TGCTAAAATT	5400
	GAAGCGATTC AGCCAGCAAC AACAACTAAA GATAATGCGA AAGAAGCAAT TGCTACGAAA	5460
15	GCGAATGAAC GTAAAACAGC AATCGCTCAA ACGCAAGACA TTACTGCTGA AGAAATTGCA	5520
	GCGGCTAATG CGGACGTAGA TAATGCTGTG ACACAAGCAA ATAGCAACAT TGAAGCTGCT	5580
20	AATAGTCAA ATGATGTAGA CCAAGCGAAA ACGACAGGTG AAAATAGTAT TGATCAAGTA	5640
	ACACCAACAG TTAATAAAAA AGCAACTGCA CGTAATGAAA TCACAGCAAT TTTAAATAAC	5700
	AAATTGCAAG AGATTCAAGc tACGCCAGAT GCAACAGATG AAGAAAAACA AGCAGCTGAT	5760
25	GCTGAAGCAA ATACTGAAAA TGGTAAAGCA AATCAAGCCA TTTCAGCAGC AACTACTAAC	5820
	GCACAAGTTG ATGAAGCTAA AGCAAATGCA GAAGCAGCGA TTAATGCGGT AACACCAAAA	5880
	GTTGTGAAGA AACAAAGCGC TAAAGATGAA ATTGATCAAT TACAAGCAAC GCAAACAAAT	5940
30	GTTATCAATA ATGATCAGAA CGCTACAACA GAAGAAAAAG AAGCAGCTAT TCAACAATTA	6000
	GCAACAGCAG TTACAGACGC GAAAAATAAT ATTACAGCTG CAACTGATGA TAATGGTGTA	6060
	GATCAGGCGA AAGACGCTGG AAAGAATTCA ATTCAAAGCA CGCAACCAGC AACAGCGGTT	6120
35	AAATCAAATG CTAAAAATGA TGTGTGACAA GCTGTGACAA CTCAAAATCA AGCAATTGAT	6180
	AATAÉAACTG GTGCTACAAC TGAAGAGAAA AATGCAGCAA AAGATTTAGT TTTAAAAGCT	6240
	AAAGAAAAAG CGTATCAAGA TATCTTAAAT GCACAAACAA CTAATGATGT TACGCAAATT	6300
40	AAAGATCAAG CAGTTGCTGA TATTCAAGGT ATTACTGCAG ATACAACAAT TAAAGATGTT	6360
	GCGAAAGATG AATTAGCAAC AAAAGCAAAC GAACAAAAAG CGCTTATTGC ACAAACTGCA	6420
45	GATGCGACTA CTGAAGAAAA AGAACAAGCA AATCAACAAG TAGACGCACA ATTAACACAA	6480
	GGTAATCAAA ATATTGAAAA TGCACAGTCA ATCGATGATG TAAACACTGC AAAAGATAAT	6540
	GCAATTCAAG CAATTGACCC AATTCAAGCA TCAACAGATG TTAACCGAA TGCAAGAGCG	6600
50	GAATTGCTAA CTGAAATGCA AAATAAAATA ACTGAAATAC TTAATAATAA TGAGACTACT	6660
	AATGAAGAAA AAGGTAACGA TATTGGACCA GTTAGAGCAG CATATGAAGA AGGTTTAAAT	6720

55

	AAAGTTCAAC AACTTCATGC AAATCCTGTT AAGAAACCAG CAGGTAAAAA AGAATTAGAT	6840
	CAAGCTGCAG CTGATAAGAA AACACAAATA GAACAAACAC CAAATGCATC ACAACAAGAA	6900
5	ATTAATGATG CAAAACAAGA AGTTGATACT GAATTAAATC AAGCGAAAAC AAATGTCGAT	6960
	CAATCATCAA CAAATGAATA TGTGATAAT GCAGTTAAAG AAGGAAAAGC TAAAATTAAT	7020
	GCAGTAAAA CATTAGTAGA GTACAAAAAA GATGCTTTAG CTAAAATTGA AGATGCATAT	7080
10	AATGCTAAAG TAAACGAAGC GGATAACTCT AACGCATCGA CTTCAAGTGA AATTGCTGAA	7140
	GCGAAACAAA AACTTGCTGA ATTAACAACTGCGGATC AAAATGTTAA TCAAGCTACT	7200
	TCTAAAGATG ACATTGAAGT TCAAATTCAT AATGACTTAG ATAATATTAA CGATTACACA	7260
15	ATTCCAACAG GTAAAAAAGA ATCAGCTACA ACAGATTTAT ATGCTTATGC AGATCAGAAG	7320
	AAAAATAATA TTTCAGCTGA CACTAATGCA ACACAAGATG AAAAGCAACA AGCAATTAAG	7380
20	CAAGTTGACC AAAATGTTCA AACTGCATTA GAAAGCATTATAATGGTGT GGATAATGGT	7440
	GACGTTGATG ATGCATTAAC ACAAGGTAAA GCAGCAATTG ATGCTATTCA AGTAGATGCT	7500
	ACTGTTAAAC CTAAAGCGAA CCAAGCTATT GAAGTTAAAG CAGAAGATAC GAAAGAATCT	7560
25	ATTGATCAAA GTGACCAGTT AACTGCTGAA GAAAAAACTG AAGCATTAGC AATGATTAAA	7620
	CAAATTACAG ATCAAGCTAA ACAAGGTATT ACTGATGCAA CAACAACCTGC TGAAGTTGAA	7680
	AAAGCGAAAg cTCaAGGACT TGAAGCATTT GATAACATTC AAATCGACTC AACAGAAAAA	7740
30	CAAAAAGCTA TCGAAGAATT AGAAACTGCA CTAGACCAGA TTGAAGCAGG TGTAATGTC	7800
	AACGCTGATG CTACAACTGA AGAAAAAGAA GCGTTTACGA ATGCTTTAGA AGACATTTTA	7860
	TCAAAAGCAA CTGaAGATAT TTCTGATCAA ACTACAAATG CAGAAATCGC TACTGTCAAA	7920
35	AATAGTGC GC TTGAACAACT TAAAGCACAA CGTATTAATC CTGAAGTTAA GAAAAATGCT	7980
	TTGGAAGCAA TCAGAGAAGT GGTAAACAAG CAAATAGGAA tAATTAAAAA TGCAGATGCA	8040
	GATGcATCGG CGGAAAGAnA TTGCACGTAC GGGATTTAGG TAGATATTTT GGACCGATTT	8100
40	GCTGGATAAA TTTAGGGTnA AACCCCAACC AATGCCGAAG TTGCCTGAAT TACCA	8155

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1630 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

	CTGTTTTATT TGCAGCACCC ATACTGGAAA TCACTTTAAT CCCTCGGTCA AGACACTCTT	120
	TCATTAAGTG TACTTTGTAC ATTATTGTAT CACTTGCATC TACAAAATAA TCTATATCGT	180
5	AGTTATCGAA AATTTCTTCA TATGTCTCTT CTGTATAAAA CATATGTAAG GGCGTGACTT	240
	TACAATCTGG ATTAATTAAT TTAATACGTT CTTCCATCAA AGAAACTTTA CTTTGTCCTA	300
	CCGTTGTAGT TAAAGCGTGT AATTGTCTGT TTACATTTGT AATATCAACA TCATCTTTAT	360
10	CTATTAATAT AATATGACCA ATATTCGTTT TTGCTAATGC TTCAGCAGCA AATGAACCAA	420
	CACCTCCAAC GCCAAGTATG ACAACAGTTT GTTGCTTCAA TAAATCTAAA CCTTGTGTGC	480
	CAATCGCTAG TTCATTTCTT GAAAATTGAT GTTTCATTAT TTTACCTCTT TCACTGATTT	540
15	ATACATAAGT ACATAGTAAC TTAAAATTTT ATATTTAGCA TTATCACTTT GATTATTTTC	600
	CCAAAATTCA ACGAGGAAAC ATTTATTAAA CGCTATAAAA CCCAACTAAT TCTTTATTAA	660
20	AAACTTAAAG AAACGCATAA AAATACGCAA GACAAAGTCT TGCGTATCGA TAGAGTCCGT	720
	ATTGCCGTAG TTATAATAGC TTGATCATTC GGCCTGTTAT ATACAGGTGG GTGCCCTGTT	780
	TCTTGTTTTG TACGTCCTTC ATATAAGGCG TGTACGCTGC AAGAAAACCC ATTGGGCTCC	840
25	CTTGATCAAA GAGTGTTAGG CCCAAATTAA AAAGCAAAC TACGAACAAC TCAGATGACT	900
	ATCTTATGAT GTTATATTAC CACATAATTA AAATTAATGA AATTATAACA AACCAAAGTT	960
	TATTGATTTT TTAAAATTTA GTGACGAATT CGCAAAGAAA GTTCTTCTAA TTGTTTATCA	1020
30	GAAACTTCAC TAGGCGCATT CGTTAATAAA CATGTAGCAG ATGCTGTTTT AGGGAATGCG	1080
	ATTGTATCTC TCAAGTTTGT TCTATTAGTC AATAACATGA CTAATCGGTC tAATCCTAAT	1140
	GCAATACCGC CATGTGGTGG TGCACCATAT TTAAATGCAT CTAGT _a AGAA GCCGAACGTG	1200
35	TCCT _g TGCTT GTTCTTTAGT AAATCCAAGA ACTTCGAACA TTTTCTTCTG TAACTCACCA	1260
	TCAT _s AATTC TGATTGAACC GCCACCTAAT TCATAACCAT TTAATACTAT GTCATAAGCA	1320
	TTTG _c CTCAG CTTCTTCTGG CGCAGTGCCA AGCTTAGCAA TATCAGCTTC TTTGGAGAT	1380
40	GTAAATGGAT GATGTGCTGC AACGTAACGT TTCGCATCTT CATCATATTC TAATAATGGC	1440
	CAATCTGTCA CCCATAAGAA GTTTAATTTT GTTTCATCGA TTAAACCTAA TTCTTTAGCT	1500
45	AATTTGACAC GTAATGCACC TAACTTTGT GCAACGACAT TTGGT _t tGTC TGCAACAAAC	1560
	ATTACTAAGT CACCAGCTTC AGCACCAGTT AATGTAAGTA ATGTTTCAAC ATTTTCTGTT	1620
	cAAAGAAACG	1630

50 (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 732 base pairs

55

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

	CAATTGGACA TCTTGTATGA AAAGGACAAC CTGCGGCGG ATTACTTGGC GAAGGTAATT	60
10	CTCCTTTTAA TATAATTCTA TTGTTATTAT GTTTATCAAT TTGTGGTATT GATGAAATCA	120
	ACGCTTTTGT ATATGGATGT TTGGGATTTT CATAAATTTT TTTATCAGAT GCGATTTCAA	180
	CTATATGACC TAAATACATA ACTCCAATGA CATCACTTAT ATGTTTTTACT ACACTTAAAT	240
15	CATGTGCGAT AAATAAATAG CTTAAGTTAA ATTGTTCTTG TAAATCTTTT AATAAATTCA	300
	GTAATTGAGA TTGAACAGAT ACATCTAATG CACTTACAGG CTCATCAGCA ACAATTAAAC	360
	TCGGACGCAA AGCCAATGCT CTGCAATTC CCACTCTTTG TCTCTGTCCA CCTGAAAATT	420
20	CATGTGCATA TTATAATAT GCATCTTCAC TTAGGCCAAC ACATTTTAAT AAATATAGTA	480
	CTTCTTTTTT TATTTCTTCT TTGGCAATT TTTTATAATT TAAATAGGT TCTGAAATGA	540
	TATCTCCAAC CATTTGCATC GGATTCAATG ATGCATACGG ATCTTGAAAT ATCATCTGAT	600
25	ATTGTTGTCG TGATTTTCTG AGTTTTTTAC CTTGTAATCT TGTTATATCT TCACCATTA	660
	CAATTATTGA GCCTGAAGTT GCATCTTCAA GCCTGATAAT CACTTTACCT AACGTTGACT	720
30	TACCACAACC CG	732

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5838 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	AATATATTCA TATGTTTCAT CAACAATATT AGCTGCTTTT TGAATTAAAG CAATTTTCGTC	60
45	AGCATCTTTG ACGTCTCTAA TTTTATCTAC AGTATTAGAA ATGCTTATTA ATGATATACG	120
	GCTTTTATTT AATTCAAGGT ATGTATCATA ACTTACATGA TGCCCCTCAA AACCTACATT	180
	TTCAAAATTT TCTTGGTGTA GCAATTCTTT AATCTCACCA ATAATAGTAG ATTTACGATT	240
50	AATAATTTCA TAATTTGGCG CCTGCTTAGT TGCTTGATCA ATATATCTAA AGTCTGTTAT	300
	CAAATATTGT TTATCTTTAG ATATGATAAG TGCTCCACTG GTACCAGTAA AACCTGATAA	360
	ATATCTTCTA TTGTAATCCG AAAGAATGaT AATCGCATCT AAATGTTTTT GTTCTAAAAT	420

55

	CAACTTTATA CATTAAAATA ATATCATAAT AAGGATAAAA AATAATAGAT ATTGATTTTA	540
	GGGAGATAGT AATGAAAAAA TTGGTTTCAA TTGTTGGCGC AACATTATTG TTAGCTGGAT	600
5	GTGGATCACA AAATTTAGCA CCATTAGAAG AnAAAAACAAC AGATTTAAGA GAAGATAATC	660
	ATCAACTCAA ACTAGATATT CAAGAACTTA ATCAACAAAT TAGTGATTCT AAATCTAAAA	720
	TTAAAGGGCT TGAAAAGGAT AAAGAAAACA GTAAAAAAC TGCATCTAAT AATACGAAAA	780
10	TTAAATTGAT GAATGTTACA TCAACATACT ACGACAAAGT TGCTAAAGCT TTGAAATCCT	840
	ATAACGATAT TGAGAAAGAT GTAAGTAAAA ACAAGGCGA TAAGAATGTT CAATCGAAAT	900
	TAAATCAAAT TTCTAATGAT ATTCAAAGTG CTCACACTTC ATACAAAGAT GCTATCGATG	960
15	GTTTATCACT TAGTGATGAT GATAAAAAAA CGTCTAAAAA TATCGATAAA TTAAACTCTG	1020
	ATTTGAATCA TGCATTTGAT GATATTAAAA ATGGCTATCA AAATAAAGAT AAAAAACAAC	1080
20	TTACAAAAGG ACAACAAGCG TTGTCAAAT TAAACTTAAA TGCAAATCA TGATAGGAGT	1140
	CTTTTAATGC GTAATATAAT ATTTTATCTT GTACTTATTA TTGCTGCGAT TGGATTAGTA	1200
	ATGAATCTAG ATGCCTTTAT TTTTCAATC GTCAGAATGT TAATCAGCTT TGcgTAaTAG	1260
25	CTGGTATTAT TTATCTGATT TATTATTTCT TCATCTTAAC TGAAGACCAA CGCAAATATC	1320
	GCAAAGCAAT GCgTrAaGTA TAAAAGAAAT CAAAGAAGAA AATAGATAAA AAAACGGAAG	1380
	CACTTGTAGG TAAATAGTC TACGTGCTTC CATTTTTTAT TCTAAAACT ACTTTCTAAA	1440
30	CATCCATTCA TCTGAACGAT ATTTTTCAGT TAATTCTTCC ACTTCTGCCA ATTGAGCTTC	1500
	TGcTAATTCA AGTGGCTTTA ATTCTATATT TAAACCTTTC TTAAACCTT TCTCGAAAGC	1560
	TTCTTCCATT TGACTAATAG TAATGTGTTT ATCTGAAATA TCATTGATGG CAACTGCTTT	1620
35	TTCAACGAAT GCCTCTTTCA TTTTAAATT TAATCTTTCA TTTTATAAA TrAACATATC	1680
	AAACAgTTCA TCAATATCAA TATCTTGTA AATCGAACCG TGTTGGAGGA TTACGCCCTT	1740
40	TTGTCTCGTT TGAGCACTCC CAGCAATCTT ACGGCCTTCA ACAACTAGCT CATACCAACT	1800
	TGGTGCATCA AAACACACTG AACTTCGAGG TTGTTTAAAT TTTTGACGCT CTTCAGGCGT	1860
	TTTAGGTACC GCAAATAAG TATCAAATCC TAAGTTTTTA AATCCTTCTA ATAATCCTTG	1920
45	TGAAATCACT CTGTACGCTT CTGTAAGTGT AGAAGGCATA TTCGGATGCG ATTCAGGCAC	1980
	AATCACACTG TAAGTTAACT CTTTATCATG TAGCACCCCA CGGCCACCAG TTTGACGCCT	2040
	TACGAGACCA AAACCTTTCT CTTTAACTT ATCAATATCA ATTTCTTTTT GTAGCCTTTG	2100
50	GAAATACCTT ATTGATAATG TTGCAGGATT CCATGTGTAA AAACGTATAA CTGGATCAAT	2160
	TTCACCTCTA GAGACAAAAT TTAATAACGC TTCATCCATT GCCATATTAT AATATGGGTC	2220

55

	AAATGTATAA	TATTTGATTC	GCTAATTAAT	CAATTTAACT	AAATGAATAA	TAATTGCAAT	2340
	TCTTTAGTGA	AATATTTTGA	TAATTTGACC	TAACAGTCTT	ATAATTATAT	TATCGTTTAA	2400
5	TTAGGGAGGA	TGCAAGATGA	GTGCTAGTTT	GTACATCGCA	ATAATTTTAG	TTATAGCAAT	2460
	TATTGCTTAT	ATGATTGTTC	AACAAATTCT	TAACAAGCGA	GCTGTTAAAG	AATTAGATCA	2520
	AAATGAATTC	CATAATGGGA	TTAGAAAAGC	TCAAGTCATC	GATGTTAGAG	AGAAAGTTGA	2580
10	CTATGACTAC	GGTCACATTA	ATGGGTCTCG	CAATATTCCT	ATGACAATGT	TCAGGCAACG	2640
	ATTCCAAGGA	TTAAGAAAAG	ATCAACCGGT	ATACTTATGT	GATGCCAATG	GGATTGCTAG	2700
	CTATAGAGCC	GCTCGTATTT	TGAAAAAGAA	TGGATATACA	GATATCTATA	TGTTAAAAGG	2760
15	CGGCTATAAA	AAATGGACTG	GAAAAATAAA	GTCTAAAAAA	TAGTTTTTGT	AAATTTAATA	2820
	TACGATTTAA	TAAAATCTGA	GTGTTAATTG	ATCATCAATA	ACAATACTCA	GATTTTAATT	2880
20	TTTTAACAAA	GTCTGTTACT	ATATTTCTCT	AGCTTCACTG	ATCATTAAAC	TTAGTTTCAG	2940
	CATAATAAAG	AAAGTTCAGC	TCATTTTCAA	TACGATTCAA	TTACCGCAAT	CTAAAAAATG	3000
	AAAAGACAAT	TTCTATGAAA	GAATAATACC	AAACCCTAAG	AGTTATTACT	TCGGTTTAGT	3060
25	TTTCTTGTTT	AAATAGAAAT	TGTCTTTTTC	AATTGATTTT	GAAACCATTA	TCCTTAAATC	3120
	TTCATACAAA	GTTAGAATAA	TAATTCCTCG	AATATGTGTT	TAATACTTTA	TTTTTCCTGT	3180
	TTAAGATTTT	CAAACTTTAA	TATTTGTTTA	CGAGCAGCTG	TAGCTTCGTC	TAATCGATCA	3240
30	ATCACAGTTG	TATGTGGTGC	TTCTAGCacT	TTATCAGGAT	CATTTTTAGC	TTCTTCAGCA	3300
	ATACTAATTA	ATGTATCGAT	AAAATAATCA	AGTGTTCCTT	TAGACTCTGT	CTCAGTCGGT	3360
35	TCAATCATCA	TACCTTCTTC	AACATTTAAT	GGGAAGTATA	TTGTTGGTGG	ATGTACACCG	3420
	AAATCTAATA	ATCGCTTAGC	CATGTCTAAA	GTACGTACAC	CAAATTCCTT	TTGACGCACA	3480
	CCACTTAACA	CAAACTCGTG	TTTACAATAT	TGTTTATAAG	GTATTTCAAA	GTGTTTAGAT	3540
40	AAACGTGCTT	TAATATAATT	CGCATTAAGA	ACCGCTGCTT	CAGAAACCTC	TTTAAGTCCA	3600
	GTTGCTCCCA	TAGTTCGAAT	ATACGTATAA	GCTCTTAAGT	AAATACCAAA	GTTACCATAA	3660
	AATGGTTTTA	CACGTCCGAT	AGAATTTTTA	ATGTCATTAT	CATATTTAAA	TTTGTGCCCA	3720
45	TCTTTAATAA	CCATTGGCTT	TGGTAAGTAA	CTTGCTAGTT	CTTTTACTAC	ACCGACTGGA	3780
	CCTGAACCAG	GACCGCCACC	ACCATGTGGA	CCAGTAAATG	TTTATGCAA	GTTTAAATGA	3840
	ACAGCATCAA	ATCCCATATC	TCCTGGGCGA	ACTTTGTCCA	TAATAGCGTT	TAAATTCGCA	3900
50	CCATCATAAT	ATAATAGACC	ACCAGCATTA	TGGACGATTT	CACGGATTTT	CATAATATTT	3960
	TTTTTCGAAAA	TACCTAAAGT	GTTTGGATTA	GTAAACATAA	TAGCTGCTGT	ATTTTCATTT	4020

55

	GATTTAAATC CTGCAATGa AGCTGAGGCT GGaTTCGTAC CATGCGCAGA ATCTGGcACA	4140
	ATGACTTCAT CACGATGACC TTCACCATTA TTCTCATGGT AAGCTTTAAA TATCATCAAT	4200
5	GCAGTCCATT CACCATGTGC GCCAGCAGCT GGTGTGAATG TCACCTCATC CATACCAGTA	4260
	ATTTCTTTTA ATTCTTCTTG CAAACTATAA ATAATTCTA ATGAACCTTG AACTTGATCT	4320
	TCATCTTGTA ATGGATGTGA TTCACTAAAT CCTGGTATTC TAGCAACCTT TTCATTAATT	4380
10	TTAGGGTTAT ACTTCATCGT ACATGAACCC AATGGATAAA ATCCGTTGTC TACACCGAAA	4440
	TTTTTATTG AAAGTTCAGT ATAATGACGT ACTAAGCTA GTTCAGCAAC TTCAGGAAAC	4500
	TCCGCTTTGT TTTTACGAAT AAATTATCA TCTAACAATG ACTCAACAGA ATTTGTTTTA	4560
15	ATATCACTTT TTGGTAATGA ATATGCATAT CTGCCTTCAC GAGATCTTTC AAAAATTAAT	4620
	GGACTTGATT TACTAGTCAT TTAACCTACC AGCCTTTTCT ACAAATGTAT CGATTTTCATC	4680
20	TTTTGTCTT AATTCACTTA CAGCTATTAA CATGTGATTT TTAAAGTCGT CTGAAACAAC	4740
	ACCTAAATCA AAACCACCGA TAATATTGTA CTTCACTAAT TCCTCGTTAA CTTGTTGAAT	4800
	TGGTTTGTCA AATTGACTA CAAACTCATT GmnAAGnTGT ACCATCTAAT ACTTCAAAAC	4860
25	CTTTTTTAAT AAATTGTTGT TTAGCATAGT TAGCATGTTT TATATTTTGA ACTGCAATAT	4920
	CATAGATACC TTGTTTACCA AGTGCTGACA TTGCAATTGA TGaCGcTAAA GCATTTAATG	4980
	CTTGGTTAGA ACAAATATTA GATGTCGCTT TATCGCGTCG AATATGTTGT TCACGTGCTT	5040
30	GTAATGTAA TACAAAGCCA CGATTACCTT CATCATCTTG TGTTTGACCG ACTAATCTAC	5100
	CTGGCACTTT ACGCATTAAC TTTTTCGTCG TTGCAAAATA TCCACAATGT GGCCACCGA	5160
	ATTGAGCAGG AATCCGAAT GGCTGAGTAT CACCTACAAC AATATCTGCA CCAAATGAAC	5220
35	CTGGAGGTGT AAGTAATCCC AATGCTAATG GATTTCATA TACGATAAAT AATGCTTTTT	5280
	TATCTFICAAT AAAGCTATGA ATCTTTTCAA GATCTTCAAT TGAACCGTAA AAGTTTGGAT	5340
40	ATTGTACTGC AACAGCTGCT GTTTCATCAT CCACTGCTGC TTCTAATTTT TTCAAATCTG	5400
	TAACAGTGCC ATCTAAATCG ATTTCCAATA CTTCGAATTC CTTACGCGTC TTAGCATAAG	5460
	TATGAAGTAC TTGTAATGCT TGATAATGTA AACCTTTTGA GACTACAATT TTATTTTTCT	5520
45	TTGTTTGAAT AAATGCTAAG ATACATGCTT CAGCAAAGCT AGTCATCCCA TCATACATAG	5580
	AAGAATTTGC TACATCCATA TCTGTTAATT CACAAATTAA AGTTTGAAC TCAAAAATGG	5640
	CTTGTAATTC ACCTGAGAA ATTTCCGGTT GATATGGCGT ATATGCTGTG TAAAATTCTG	5700
50	ATCTTGAAAT CATAGCATCC ACAACTGATG GCGCGTAATG ATCATAAACA CCAGCACCCA	5760
	rAAATGATGT ATGCGTTTCT TTAGTGATAT tCTTGCTkGC AATGGGGATT TAAACnTCTA	5820

55

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18355 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

ATrATAATTG GCTTTGCTAA TAATTACTTC CCTGAATTAC aAGTATTAGC AAACGAAATA 60
 AAATCTGATA TGGCTAGTTC ATTA AAAACAA TGATATTTT ATTAAATTT TTaAGCTTT 120
 GTACGAAATT GTACAAAGCT TTTTGGTGC GTATTGTATG GGCAACAACT TGACGATGAA 180
 AATCCGTTAC AGGATTGGTA ATAGGAAATG TTAGCGAAAG ACAAGGGTAT CCATTGTAGA 240
 TTAACAAAAG GACGTTTCCA CAAGTGTGGG TTATTCTCAC TAAAGCAATA CGCAGAGACA 300
 ACTTACGTAA AATTTTGAAC TGA CTAGAAC GGAAC TTCTA CTCAATTATT GATAAAAATT 360
 TTCAAAAAGA CTTGAATGTG CTGAGAATAC GAAGTTTATG GAAGGATTAT CAAAATATAA 420
 ATGTGCATTC ATTTACAACC TTTATTGACA ATGATTCTCA ACTAATATAG TATATAATCA 480
 AATCGTAATA GTTACGATT GTTTCTGCA ACTTTTGA AGTTT TAGTT GAGGTGAAAA 540
 CAATAAAAGC ATCTAAGTGA ATGTAGTTAA CGGACAAC TG CATTGCTTG TAGAGCCACA 600
 AGAAGCAACT TTAAATAAGG TTTACGGTTG CATT TTGATA CAACAACCGA T TACTAAGTC 660
 ATGCTTTCCA CTTTGC GGGT TAGCATGACT TACCTAATAG ATAGAGCTAT TAGGTT CAGC 720
 TTCTAAAAA TTACAGTTT AGAGGAATAC AGTTGcTTGc tTCGCAACAA CTGCATAAGA 780
 GCCATGGTTT TCGCTTTTGC GAATTAGCAT GACTTACCTA CTAGATAGAG CTATTAGGTT 840
 CATCTTCTAA AAAATTACAG GTTTAGAGGA ATACAGTTGT TTGcTTGCA ACAACTGCAT 900
 AAGAGCCTCT AGTAATTAAA ATTACAGAGG CTCTAAAAAT ACATCTAAAG GAGTGTGCGTA 960
 TGAATCGGCA GGTTATAGAA TTTTCTAAGT ATAATCCTTC GGGGAATATG ACGATACTTG 1020
 TTCATTCAA ACATGATGCT AGTGAATATG CATCTATCGC CAATCAGTTG ATGGCCGCAA 1080
 CACATGTATG CTGTGAACAG GTAGGCTTTA TAGrATCAAC ACAAATGAT GATGGTAATG 1140
 ATTTTCACTT AGTTATGAGC GGTAATGAAT TTTGCGGTAA TGCGACGATG TCATATATAC 1200
 ATCATTGCA GGAAAGTCAT TTGCTTAAAG ACCAACAGTT TAAGGTGAAG GTGTCTGGCT 1260
 GTTCGGATTT AGTGCAATGC GCAATTCATG ATTGCCAATA CTATGAAGTT CAAATGCCAC 1320
 AAGCCCATCG TGTTGTGCCA ACAACAATTA ATATGGGTAA TCATT CATGG AAAGCAATAG 1380

	TTCAACATTT	GGTTGAAGCG	TTTGTGCGTG	AgcAACAATG	GAGTCACAAA	TATAAACAG	1500
	TAGGTATGAT	GCTTTTGTGAT	GAACAACGTC	AATTTTTACA	GCCATTAATC	TATATACCAG	1560
5	AAATTCAAAG	TTTAATTTGG	GAAAATAGCT	GTGGTTCTGG	TACAgcATCA	ATTGGGGTTT	1620
	TTAATAATTA	TCAACGTAAT	GACGCATGCA	AAGATTTTAC	AGTACATCAG	CCAGGGGGCA	1680
	GTATTTTAGT	GACATCAAAG	CGATGTCATC	AATTGGGATA	TCAAACCTCA	ATTAAAGGAC	1740
10	AGGTTACAAC	TGTAGCTACA	GGaAAAGCAT	ATATAGAATA	AGGAGCCTAC	AATGAATAAC	1800
	TTTAATAATG	AAATCAAATT	GATATTACAA	CAATATTTAG	AAAAGTTTGA	AGCGCATTAC	1860
	GAGCGTGTAT	TACAAGACGA	TCAATATATC	GAAGCATTAG	AAACATTGAT	GGATGACTAT	1920
15	AGTGAATTTA	TTTTAAATCC	TATTTATGAA	CAACAATTTA	ATGCTTGGCG	TGACGTTGAA	1980
	GAAAAAGCAC	AATTaATAAA	ATCACTGCAA	TATATTACAG	CGCAGTGTGT	TAAACAAGTG	2040
20	GAAGTCATTA	GAGCGAGACG	TCTATTAGAC	GGACAGGCGT	CTACCACAGG	TTACTTTGAC	2100
	AATATAGAAC	ATTGTATTGA	TGAAGAGTTT	GGACAATGTA	GTATAGCTAG	CAATGACAAA	2160
	TTATTGTTAG	TTGGTTCAGG	TGCATATCCA	ATGACGTTAA	TTCAAGTAGC	AAAAGAAACA	2220
25	GGTGCTTCAG	TTATCGGTAT	TGATATTGAT	CCACAAGCCG	TTGACCTAGG	GCGCAGAATC	2280
	GTTAACGTCT	TAGCACCAAA	TGAAGATATA	ACAATTACGG	ATCAAAAGGT	ATCTGAACTT	2340
	AAAGATATCA	AAGATGTGAC	GCATATCATA	TTCAGCTCGA	CAATTCCTTT	AAAGTACAGC	2400
30	ATTTTAGAAG	AATTATATGA	TTTAACAAAT	GAAAATGTCG	TAGTTGCAAT	GCGCTTTGGT	2460
	GATGGCATCA	AAGCAATATT	TAATTATCCG	TCACAAGAAA	CAGCGGAAGA	TAAGTGGCAA	2520
	TGTGTGAATA	AACATATGAG	ACCACAGCAA	ATTTTTGATA	TAGCACTTTA	TAAAAAAGCA	2580
35	GCTATAAAGG	TAGGTATTAC	GGATGTCTAA	ATTATTAATG	ATAGGCACTG	GTCCgGTCCG	2640
	AATGCAATTA	GCGAATATTT	GCTATTTAAA	ATCAGATTAT	GAGATTGATA	TGGTTGGACG	2700
40	TGCCTCAACA	TCAGAAAAAT	CAAAACGCTT	ATATCAAGCG	TATAAAAAAG	AGAAACAATT	2760
	TGAAGTCAAA	ATACAAAACG	AGGCGCATCA	ACATCTGGAA	GGTAAGTTTG	AAATTAATCG	2820
	TTTGATATAA	GATGTTAAAA	ACGTTAAGGG	TGAATACGAA	ACGGTTGTCA	TGGCATGCAC	2880
45	AGCAGATGCT	TATTATGACA	CACTACAGCA	ATTGTCGTTA	GAACTTTTGC	AAAGTGTCAA	2940
	ACATGTCATT	TTAATATCAC	CGACATTTGG	TTTCGCAAATG	ATTGTCGAAC	AATTTATGTC	3000
	TAAATTTAAT	AAAGATATCG	AAGTGATTTT	ATTCTCAACT	TATCTTGGCG	ATACACGTAT	3060
50	TGTTGATAAA	GAAGCGCCTA	ATCATGTGTT	GACAACAGGT	GTAAAAAAGA	AATTGTACAT	3120
	GGGATCGACA	CATTCAAAC	CAACAATGTG	TCAACGAATC	TCTGCTTTAG	CTGAGCAATT	3180
55							

	TTATGTGCAC CCACCACTAT TTATGAATGA CTTTTCATTG AAAGCCATTT TCGAAGGAAC	3300
	AGATGTACCG GTTTATGTGT ATAAGTTATT TCCTGAAGGA CCGATAACGA TGACACTAAT	3360
5	CCGTGAAATG CGTTTAATGT GGAAGGAAAT GATGGTTATT TTACAAGCAT TTAGAGTGCC	3420
	GTCAGTCAAC CTGCTTCAAT TTATGGTGAA GGAAAATTAT CCAGTACGTC CTGAAACTTT	3480
	GGATGAAGGT GATATTGAGC ATTTTCGAAAT CTTGCCAGAT ATCTTACAAG AATATCTGCT	3540
10	TTATGTAAGA TATACCGCAA TCCTCATTGA TCCATTTTCA CAGCCAGACG AAAACGGACA	3600
	TTACTTTGAT TTTTCAGCTG TACCATTTAA GCAAGTCTAT AAAAATGAAC AGGATGTTGT	3660
	TCAAATTCCA AGAATGCCAA GTGAAGATTA TTACAGAACG GCGATGATTC AGCATATTGG	3720
15	GAAAATGCTA GGTATCAAAA CGCCAATGAT TGATCAGTTC CTAAGTCGCT ATGAAGCAAG	3780
	TTGCCAGGCG TACAAGGATA TGCATCAAGA TCAACACTTA TCTTCTCAAT TTAATACAAA	3840
20	TCTATTTGAA GGAGATAAAG CACTCGTCAC AAAATTTTGT GAAATCAATA GAACGCTTTC	3900
	ATAATAAGGG TTTGAAGTTT TATAATAGAA AAAAATTATT GAATTATGTT TGACATTTAC	3960
	ATAAAAATAA GCAAATAATT GAGAAAAATA ATCATTACGA TTTGATTAAG TAATGCAACT	4020
25	TATCAATTTA GAAAGAGGAA AAGCAAATGA GAAACTAAC TAAAATGAGT GCAATGTTAC	4080
	TTGCATCAGG GCTAATTTTA ACTGGTTGTG GCGGTAATAA AGGTTTAGAG GAGAAAAAAG	4140
	AAAACAAGCA ATTAACGTAT ACGACGTTA AAGATATCGG TGATATGAAT CCGCATGTTT	4200
30	ACGGTGGATC AATGTCTGCT GAAAGTATGA TATACGAGCC GCTTGTACGT AACACGAAAG	4260
	ATGGTATTAA GCCTTTACTA GCTAAAAAGT GGGATGTGTC TGAAGATGGG AAGACATACA	4320
	CGTTCCATTT GAGAGATGAC GTTAAATTCC ATGATGGTAC GCCATTTGca TGctGACGCA	4380
35	GTTAAGAAAA ATATTGACGC AgTTCAAGAA AACAAAAAAT TGCATTCTTG GTTAAAGATT	4440
	TCGACATTAA TTGACAATGT TAAAGTTAAA GATAAGTACA CGGTTGAATT GAATTTGAAA	4500
40	GAAGCATATC AACCTGCATT GGCTGAATTA GCGATGCCTC GTCCATATGT ATTTGTGTCT	4560
	CCAAAAGACT TTaAAAACGG TACAaCAAAA GATGGCGTTA AAAAGTTCGA TGGTACTGGT	4620
	CCATTTAAAT TAGGTGAACA CAAAAAGAT GAGTCTGCAG ACTTTAACAA AAATGATCAA	4680
45	TACTGGGGCG AAAAGTCTAA ACTTAACAAA GTACAAGCAA AAGTAATGCC TGCTGGTGAA	4740
	ACAGCATTCC TATCAATGAA AAAAGGTGAA ACGAACTTTG CCTTCACAGA TGATAGAGGT	4800
	ACAGATAGCT TAGACAAAGA CTCTTTAAAA CAATTGAAAG ATACAGGTGA CTATCAAGTT	4860
50	AAGCGTAGTC AACCTATGAA TACGAAAATG TTAGTTGTCA ATTCTGGTAA AAAAGATAAC	4920
	GCTGTGAGTG ACAAACAGT CAGACAAGCG ATTGGTCATA TGGTAAACAG AGATAAAATT	4980

55

	ACAGACATTA ATTTTCGATAT GCCAACACGT AAGTATGACC TTAAAAAAGC AGAATCATT	5100
	TTAGATGAAG CTGGTTGGAA GAAAGGTAAA GACAGCGATG TTCGTCAAAA AGATGGTAAA	5160
5	AACCTTGAAA TGGCAATGTA CTATGACAAA GGTTCCTCAA GTCAAAAAGA ACAAGCAGAA	5220
	TACTTACAAG CAGAATTAA GAAAAATGGGT ATTAAGTTAA ACATCAATGG CGAAACATCA	5280
	GATAAAATTG CTGAACGTCG TACTTCTGGT GATTATGACT TAATGTTCAA CCAAACCTGG	5340
10	GGATTATTGT ACGATCCACA AAGTACTATT GCAGCATTTA AAGAGAAAAA TGGTTATGAA	5400
	AGTGCAACAT CAGGCATTGA GAACAAAGAT AAAATATACA ACAGCATTGA TGACGCATTT	5460
	AAAATCCAAA ACGGTAAAGA GCGTTCAGAC GCTTATAAAA ACATTTTGAA ACAAATTGAT	5520
15	GATGAAGGTA TCTTTATCCC TATTTACAC GGTAGTATGA CAGTTGTTGC ACCaAAAGAT	5580
	TTAGAAAAAG TATCATTCAC ACAATCACAG TATGAATTAC CATTCAATGA AATGCAGTAT	5640
20	AAATAAAGGA GCAATTAGAT GTTCAAATTT ATCTTAAAC GTATTGCGCT CATGTTTCCA	5700
	TTGATGATTG TAGTAAGTTT TATGACATTT CTATTGACGT ATATTACAAA TGAAAATCCA	5760
	GCTGTGACAA TTTTACATGC ACAAGGGACG CCAATGTAA CACCAGAGTT GATTGCAGAA	5820
25	ACGAATGAGA AGTACGGTTT CAATGATCCA TTATTAATTC AATATAAAAA TTGGTTACTT	5880
	GAAGCGATGC AATTTAATTT TGGTACAAGC TACATTACAG GTGACCCAGT TGCTGAACGT	5940
	ATTGGTCCAG CATTTATGAA TACATTGAAA TTAACAATAA TTTCAAGTGT TATGGTGATG	6000
30	ATTACATCAA TTATTTTAGG TGTAGTTAGT GCATTAAAAA GAGGAAAGTT CACTGATCGT	6060
	GCGATACGTT CAGTGGCTTT CTTTCTAACT GCATTACCAT CATATTGGAT AGCTTCAATA	6120
	CTTATTATTT ACGTTTCAGT GAAGTTAAAC ATATTGCCGA CTTCTGGATT AACAGGTCCA	6180
35	GAAAGTTACA TATTGCCAGT GATCGTTATT ACGATTGCCT ATGCTGGTAT TTACTTTAGA	6240
	AATGTTAGAC GCTCGATGGT GGAACAATTA AATGAAGATT ATGTACTTTA TTTAAGAGCA	6300
40	AGCGGTGTGA AATCTATCAC ATTAATGTTG CATGTGTTGC GTAATGCTTT ACAAGTTGCG	6360
	GTATCAATCT TTTGTATGTC TATACCAATG ATAATGGGTG GACTAGTTGT TATCGAGTAT	6420
	ATCTTTGCAT GGCCTGGACT AGGTCAATTA AGTTTAAAAG CAATACTTGA ACACGATTTT	6480
45	CCAGTCATTC AAGCATATGT ATTAATTGTA GCGGTATTAT TTATTGTATT TAATACATTA	6540
	GCAGATATCA TTAATGCGCT ATTAAATCCA AGATTAAGGG aGGGCGCACG ATGATAATTT	6600
	TAAAmCGATT ATTMCarGwT AAAGGTGCAG TAATTGCTTT AGGCATTATT GTATTATATG	6660
50	TCTTTTTAGG ATTAGCAGCA CCACTTGTA CATTTTATGA TCCTAACCAT ATCGATACAG	6720
	CAAACAAATT TGCTGGCATG AGTTTTCAAC ATCTACTAGG TACTGACCAT TTAGGTAGAG	6780

55

	TATTTGTTTC TGTACTTATT GGATCTATTT TAGGATTCTT ATCAGGATAT TTCCAAGGGT	6900
	TTGTTGACGC CTTAATCATG CGTGCGTGTG ATGTTATGTT GGCATTCCCA AGTTATGTTG	6960
5	TAACGTTAGC ATTAATTGCA TTGTTTGGAA TGGGTGCCGA AAATATTATC ATGGCATTTA	7020
	TTTTGACGCG TTGGGCATGG TTCTGTCGTG TTATACGTAC AAGTGTTATG CAGTACACTG	7080
	CTTCTGACCA TGTAAGATTT GCTAAAACAA TCGGTATGAA TGATATGAAA ATTATTACCA	7140
10	AACATATTAT GCCATTAACA TTAGCAGATA TTGCTATCAT CTCTAGTAGC TCGATGTGTT	7200
	CAATGATCTT GCAAATATCT GGCTTTTCAT TTTTAGGATT AGGTGTCAA GCGCCTACTG	7260
	CAGAGTGGGG CATGATGCTT AACGAaGCTA GAAAAGTGAT GTTTACACAT CCTGAAATGA	7320
15	TGTTTGCGCC AGGTATTGCC ATAGTGATTA TAGTGATGGC ATTTAACTTC TTATCCGATG	7380
	CTTTACAAAT TGCTATTGAT CCCCGCATCT CTCTAAAGA TAAACTTCGT TCTGTGAAAA	7440
	AAGGAGTGGT GCAATCATGA CATTGTAAAC AGTTAAACAT TTGACGATTA CAGATACCTG	7500
20	GACAGATCAA CCACTCGTGA GTGATGTGAA TTTTACATTA ACTAAGGGTG AAaCTTTAGG	7560
	CGTTATTGGA GAAAGTGGTA GTGGTAAATC AATCACTTGT AAATCGATTA TTGGTTTGAA	7620
25	TCCCGAACGA CTCGGGGTGA CAGGTGAAAT TATCTTTGAT GGTACaCAA TGTTGTCATT	7680
	ATCTGAATCG CAATTGAAAA AGTACCGTGG TAAAGACATT GCGATGGTCA TGCAACAAGG	7740
	TAGTCGTGCC TTTGACCCAT CAACTACTGT CGGTAAACAA ATGTTTGAGA CTATGAAAGT	7800
30	ACATACGTCA ATGTCTACAC AAGAAATTGA AAAGACATTG ATTGAATATA TGGATTATTT	7860
	AAGTTTGAAA GATCCTAAAC GTATATTAAA ATCATACCCT TACATGTTAT CAGGAGGAAT	7920
	GTTACAGCGA TTGATGATTG CTTTAGCGTT AgcTTTgAAA CCAAAGTTAA TCATTGCTGA	7980
35	TGAGCCGACA ACGGCTTTAG ATACAATTAC ACAATATGAT GTACTGGAAG CATTATAGAA	8040
	TATTAAAAAA CACTTTGACT GTGCGATGAT TTTCAATTCA CATGATTTAA CGGTTATTAA	8100
	CAAGATTGCA GACCGTGTTG TTGTGATGAA AAATGGTCAG CTTATTGAAC AAGGGACACG	8160
40	TGAATCAGTC TTGCATCATC CAGAACATGT TTATACGArt ATTKtATTAT CAACGAAGAA	8220
	GAAGATTAAT GATCATTTTA AACATGTGAT GAGGGGTGAT GTACATGATT AAAATTAAAG	8280
45	ATGTTGAAAA GTCATATCAA AGCGCACATG TTTTAAAGCG TCGTCGAACA CCTATCGTGA	8340
	AAGGTGTGTC ATTTGAGTGT CCAATCGGTG CGACGATTGC GATTATCGGA GAAAGTGGTA	8400
	GCGGTAAATC GACGTTGAGT CktATGATAT TAGGTATTGA GAAACCGGAT AAAGGTTGTG	8460
50	TAACCTTAAA TGATCAACCG ATGCATAAGA AGAAAGTGAG ACGTCATCAA ATTGGTGCTG	8520
	TATTTCAAGA TTATACGTCA TCATTACATC CATTTAGAC TGTTAGAGAA ATCTTATTTG	8580

55

	TGTTGGAAGA AGTCGGTCTA TCTAAGGCAT ACATGGATAA ATATCCTAAT ATGTTATCAG	8700
	GTGGAGAGGC GCAACGTGTT GCGATTGCGC GTGCAATATG TATTAACCCT AAATATATTT	8760
5	TGTTTGATGA AGCCATTAGT TCACTCGACA TGTCAATTCA AACACAAATA TTAGATTTAT	8820
	TGATTCATTT ACGTGAAACG CGTCAGTTGA GTTATATTTT TATCACACAT GATATTCAAG	8880
	CTGCCACGTA TTTATGTGAT CAATTAATTA TTTTAAATA CGGAAAAATA GAAGAACAA	8940
10	TTCCGACAAG CGCATTGCAT AAAAGTGACA ATGCTTATAC AAGAGAATTA ATAGAAAAAC	9000
	AACTATCATT CTAAGGAGTG AGATAATGAA AGGTGCAATG GCTTGGCCCT TTTTGAGATT	9060
	ATATATATTA ACATTGATGT TCTTTAGTGC CAATGCAATC TTAAACGTGT TTATACCTTT	9120
15	ACGAGGGCAT GATTTAGGCG CAACGAATAC GGTATATCGGT ATCGTTATGG GGGCATACAT	9180
	GTAAACAGCA ATGGTATTTT GACCATGGGC AGGACAAATT ATTGCTCGTG TCGGTCCCAT	9240
20	TAAAGTATTA AGAATTATTT TGATTATCAA TGCCATAGCT TTAATTATTT ATGGTTTTAC	9300
	TGGCTTAGAA GGTATTTTCG TAGCACGTGT TATGCAAGGT GTGTGTACGG CATTCTTTTC	9360
	TATGTCTTTA CAGCTAGGTA TTATTGATGC ATTACCAGAG GAACATCGTT CTGAAGGTGT	9420
25	ATCATTTGTAC TCGCTATTTT CAACGATTCC AAACCTAATC GGACCATTAG TTGCCGTAGG	9480
	TATTTGGAAT GCAATAATA TTTCACTATT TGCAATTGTC ATTATCTTTA TCGCATTAAAC	9540
	AACAACATTC TTGATATCG CGTGACCTTT GCTGAACAGG AACCCGATAC GTCAGATAAG	9600
30	ATTGAAAAAA TGCCGTTTAA CGCTGTAAC TTTTTGCGC AATTTTTCAA AAATAAAGAG	9660
	TTGTTGAACA GTGGTATTAT CATGATTGTT GCATCGATTG TATTTGGTGC AGTTAGTACA	9720
	TTTGTACCGT TATACACAGT GAGTTTAGGA TTCGCGAATG CGGGAATCTT TTTGACAATA	9780
35	CAGGCCATCG CAGTTGTTGC GGCAAGATTT TACTTAAGGA AATACATTCC GTCAGATGGT	9840
	ATGTGGCATC CTAAATATAT GGTATCTGTA CTATCATTAT TAGTAATCGC GTCATTTGTA	9900
40	GTGGCATTG GTCCGCAAGT AGGTGCAATT ATTTTCTATG GTAGTGCAT ATTAATAGGA	9960
	ATGACGCAAG CAATGGTGTA CCCAACATTA ACATCATACT TAAGCTTCGT CTTACCAAAA	10020
	GTAGGTCGTA ATATGTTGTT AGGTTTATTT ATTGCCTGTG CAGACTTAGG TATATCGTTA	10080
45	GGTGGCGCAT TGATGGGACC TATTTCCGAT TTAGTAGGAT TTAAATGGAT GTATCTAATT	10140
	TGTGGTATGT TAGTCATTGT AATAATGATT ATGAGTTTCT TGAAAAAGCC AACACCACGT	10200
	CCAGCGAGTA GTCTTTAATG AAGTGAATTA AAGCATATTA AGTTAATGAA TATTTAAATT	10260
50	TTAAAAGGTA TATTGACAT GGCGATTCAT GTGCTTCATG CTAGGACATG AAACATTCTA	10320
	TATGGCTCGT TTTTAGAACG ACATATATCT AAATAAAGCA CGCTTAAG TGAGTTTGA	10380

55

	TTACATGAAA ATATGCAAAA CGAGTATAAC TGCTAATTGA TAGAAATAGC TCACCATAAA	10500
	ATTACGGTAT GATTTTAAAT ATAAGTAAGT CGCACTACCT GCTAGTATCA ATGCTGGAAT	10560
5	GAATTCCCAC CATGTATTAA TGTATGGATA GTAGAACAGA GTTCAAGGA TAATGGACAA	10620
	TACTATTGTA ATCTTTAAAG GTATTAATCT GCTTAATTCT TGAATTAAAA TATGACGGAA	10680
	AATAAGTTGA CAAATCAAAG TATTTAATAT AATGGTTAAC GAAAATATAG CTATTAAACT	10740
10	GATGGAaCCA TACCCTTTAA TGAGCGGGTA AATGTCAAAG ACAGTAAAGG AATCTACATT	10800
	TAGTGCgAAA ATATTGAAAT GATTTAAAAG TAAAAAGAGT ACGACACTTA GTGTAAATGA	10860
	TATAAGAATA TGCCATTAT ATTTAGCACT AGCAACGATT TGCGAACGTA TCATTGGAAT	10920
15	AAACGCATCT TCATGCATCA GACGAAAAAT AGCTAGTGAA ATAATAACTG CGAGTAAATA	10980
	GCTAATGTTT ATTGAAATAG GAAAAGAGAA ACCCCACGGA GCTTGTTGAG TGAATACAGC	11040
20	TACTAACCCA AAAGTTAAAA AGACGATAAT GATCGGCAAG ATGTTAACCA AAAATATGTA	11100
	AAGGAAAATA AATCCAATAT CACGTTTGAA AAAACGCGAT TGTTGGTAG CGTATTCTTC	11160
	TTCTATGTAA TGTTTATTG TATTGACAT AGTATACCTC TTAAATAGTT GTATTATATA	11220
25	GATACTTTAG CACATATTAC TTTGTATTGT ATGTTTTATA CATTAAAATT TAAAATGAAA	11280
	AACATATCAT AAAATTGTTT TATAAAATGA AGCGCTTCCA TTGTGTTTTG TTTTGTAAAG	11340
	TGTATCATAA ATATTGAATT GAAATTTTGG GGGGAGGTAT TGTAATGACG TTTCTTACAG	11400
30	TCATGCAATT TATAGTTAAC ATTATCGTTG TAGGATTCAT GCTTACGGTT ATTGTTATCG	11460
	GGCTTATTG GTTAATTAAA GATAAAAGAC AATCACAACA TAGTGATTA AGGAATTATC	11520
	CTTTACTAGC ACGTATTAGA TATATTTTCTG AAAAAATGGG ACCGGAATTA CGTCAGTATT	11580
35	TATTTTCTGG GGATAATGAA GGGAAACCTT TTTCACGTAA TGATTATAAA AATATCGTTT	11640
	TGGCTGGAAA ATATAACTCT CGTATGACCA GCTTCGGTAC TACTAAAGAT TATCAAGACG	11700
40	GCTTTTACAT ACAGAACACA ATGTTTCCGA TGCAACGTAA TGAGATTTC AGTAGATAATA	11760
	CAACATTGTT ATCAACATTC ATTTATAAAA TCGCGAATGA GCGTTTATTT AGTCGTGAAG	11820
	AATATCGTGT GCCGACAAAG ATTGATCCGT ATTACTTAAG TGATGACCAT GCAATAAAAAT	11880
45	TAGGTGAACA TTAAAAACAT CCATTTATTT TAAAACGTAT CGTAGGACAA TCTGGTATGA	11940
	GTTATGGCGC TTTAGGAAAA AATGCCATTA CAGCTTTATC TAAAGGTCTA GCTAAAGCGG	12000
	GCACTTGGAT GAATACAGGT GAAGGTGGCT TATCAGAATA TCATTTAAAA GGTAATGGGG	12060
50	ATATCATTTT CCAAATTGGT CCCGGTTTAT TTGGTGTTTC TGATAAAGAA GGTAATTTTA	12120
	GTGAAGGTTT ATTTAAAGAG GTTGACACAGT TATCTAACGT ACGCGCATTT GAGCTGAAGT	12180

55

	TTGCTAAAAT CCGAAATGTT GAACCTTATA AAACAATCAA TTCACCTAAC CGTTACGAAT	12300
	TTATTCATAA TGCTGAAGAT TTGATTCGTT TCGTCGATCA GTTGCAGCAA TTAGGTCAAA	12360
5	AACCAGTAGG ATTCAAAATT GTAGTAAGCA AAGTTTCAGA AATTGAAACA CTTGTACGTA	12420
	CGATGGTGGA ACTAGATAAG TATCCAAGCT TTATTACGAT TGATGGTGGT GAAGGTGGTA	12480
	CTGGTGCAAC ATTCCAAGAA TTACAAGATG GTGTTGGCTT ACCGCTATTT ACAGCTCTAC	12540
10	CTATTGTGTC TGGCATGTTA GAAAAATATG GTATTTCGAGA TAAAGTGAAA TTGGCGGCAT	12600
	CTGGTAAGTT AGTGACACCA GATAAAATTG CGATTGCACT AGGTTTAGGT GCAGATTTTG	12660
	TAAATATCGC ACGTGGGATG ATGATTAGTG TCGGTTGTAT AATGAGTCAA CAATGTCACA	12720
15	TGAATACGTG TCCTGTAGGT GTTGCAACGA CAGATGCGAA GAAAGAAAAA GCATTGATTG	12780
	TTGGAGAAAA GCAATATCGT GTCACAAACT ATGTAACAAG TTTGCATGAA GGCTTATTCA	12840
20	ATATTGCAGC AGCTGTTGGC GTATCCAGTC CTACAGAAAT TACTGCTGAT CATATTGTAT	12900
	ATCGAAAAGT CGATGGTGAG TTACAAACGA TACATGATTA TAAATTAAAA CTCATTAGTT	12960
	AACCTAATTA TTTTCGGGAA TTGAAAGCAG CGGATTTTAG CGTTACTGCA AATAATTTTA	13020
25	TATTAGTAGT GGATGCTGGT CACACAAGAA CTTCAAATAT TAAAGCCCTC AGAATATGAA	13080
	TTAAGGTTTG TAACCTTAGT CTTATCTGAG GGCATTTTTA AGTTATAAAC TATTTGTCGT	13140
	CCATTTTATC TTTTCTTTT AAACCTCTGT GCTTTAATTG CTTTTCAAGT TTTTCAAAAC	13200
30	TAATATCTTT ATTTTCTTTA GTCGAAACAC CAAGACGTTT ATTTAATTTT TTCATGTCAA	13260
	CTTCTGTGTA ATCTATGTCT AAGTGyTCAA TTGCTTTTTT ATCTTTATAG TCTACTTTGT	13320
	ATTTTACGCC TTAAAGGTCT TTGAAAATAC TTTTCAGATT GGCGAATAAC TTTTGGCTT	13380
35	CGTCTTTATC CATACTAGA TCGTCATATT TAATTGTGTT GATTGTAGAC TGTTTTAAAA	13440
	CTTTATCATC TTTATATGTG ATAGAAGTTA GTACATGTTT ACCACTAACA TCACCWTCAT	13500
40	ATGTTTTGGT TTGTTCTTTA CCACAAGCTG ATAATGCAAT GATACAACT AATGCTACTA	13560
	CAATTAATGA ACATAATTTT TTCAAAGTCA GTCGCCTTCT TTCGATATTT GTATTATAAA	13620
	GAAATTATAA CATTTACTAA AAAATGATGT TATTCAAAAA TTTAAATTTT GTCATTTTTT	13680
45	TTGAAGATAT GAGTTTTTTT AAGCGGATTC CTCACAAAAT TTTAAAAATA TTTAAGCCTk	13740
	AAAATGATAA AGCGkTAGGG AACGTTTTTC TGAAAGTTAG TGATACAATA GTTTTAAGTT	13800
	GAAATACAGG AGGATGAATA ACATGAATCA GTCAGTCAA TTACTTAAAC ATTTAACAGA	13860
50	TGTAAACGGC ATTGCTGGTT ATGAAATGCA AGTTAAAGAA GCAATGCGTa ACTATATAGA	13920
	GCCTGTCAGT GATCAAATTA TTGAAGATAA CTTGGGTGGC ATTTTGGAA AGAAAAATGC	13980
55		

	AACAAAGATT GATAAACATG GTTTTATTTC ATTTACGCCA kTgGTGGATG GTGGAATCAA	14100
	GTCATGCTAT CTCAAAAAGT AACGATTACA ACAGATTCGG GCAAAGAAAT TAGAGGTATC	14160
5	ATCGGTTCTA AACCGCCACA TGTCTTAACG CCTGAAGAAC GTAAAAAGCC AATGGAAATC	14220
	AAAAATATGT TTATAGATAT TGGTGTTAGT AGCAAGGAAG AAGCTGAAGA AGCTGGCGTT	14280
	GAAGTAGGCA ATATGGTTAC GCCATATAGT GAATTTGAAG TGCTTGCAA TGATAAATAT	14340
10	TTAACTGCGA AxCATTTGAT AATCGCTATG GCTGTGCATT AGCTATTGAG GTATTAAAC	14400
	GTTTAAAGA TGAAATATT GGCATTAACT TATACAGTGG TGCCACAGTG CAAGAAGAAG	14460
	TTGGTTTTCG TGGTGCGAAA GTGGCAGCGA ATACGATTAA ACCAGACTTG GCGATAgcTG	14520
15	TcGATGTAGG TATTGCTTAT GATACCCAG GTATGTCAGG TCAAACGAGC GATAGTAAAC	14580
	TAGGCGGTGG TCCAGTTGTC ATTATGATGG ATGCTACAAG TATTGCTCAC CAAGGTTTGC	14640
20	GAAAgcATaT TAAAGATGTA GCTAAGGAAC ATAACATCGA AGTACAATGG GATACGACAC	14700
	CAGGTGGAGG TACAGATGCG GGAAGTATTC ATGTCGCAA TGAAGGTATT CCAACGATGA	14760
	CAATCGGTGT TACGCTGCGA TACATGCATT CTAATGTTTC AGTGCTCAAT GTAGATGATT	14820
25	ATGAAAATTC TATCCGTCTT GTTACTGAAA TTGTCCGTTT ATTGAATGAT GAAAGTTATA	14880
	AAAATATCAT GTGGTAATCA AATCCATAAA TAATAAGAA TCCTTTTAAT ATGGTAGGTT	14940
	GTTAAACAAT TGTCTAATTT TAATTCCTAG TCATTAGACA GTATCCATGT TAATAGGATT	15000
30	TTTTGTTTTT AATTTAAATG CTGAAAATCA ATTATGCCTA AATTTTGATA TTACAAGAAA	15060
	ATGATTTTTT CTAAATGTA ATTGCACTAA AAACCAAAAA AACGGGAATA ATATACCTGA	15120
	TATATTACAT GAGGAGCGGT GCAAATGTTG TTAGAAATTA AAGATTTAGT GTATAAAGCG	15180
35	AGCGATAGAA TCATACTAGA TCATATCAGT CTAAAAGTAG ATAAAGGCGA GAGTATTGCC	15240
	ATTATAGGTC CATCAGGTAG TGGTAAAAGT ACATTTCAA AGCAAATATG TAATTTGTTT	15300
	AGTCCAAC TA GTGGAGAACT TTATTTTAAA GGTAACCCT ATAATGATTA TGACCCGGAA	15360
40	GAATTGCGTC AACGAATCAG TTATTTGATG CAGCAAAGTG ACTTGTTTGG TGAAACGATT	15420
	GAAGATAACA TGATATTCCC ATCACTTGCA CGTAATGATA AATTGATAG AAAACGTGCA	15480
45	AAGCAATTAA TTAAAGATGT CGGTTTGGGA CATTATCAAT TAAGTTCGGA AGTGAAAAT	15540
	ATGTCGGGTG GTGAGCGGCA AAGAATTGCT ATAGCGCGCC AACTGATGTA TACACCGGAT	15600
	ATTCTTTTAT TAGATGAATC GACCAGTGCA TTAGACGTTA ATAATAAGA AAAGATAGAA	15660
50	AATATCATT TTAATTAGC AGATCAAGGC GTGGCAATTA TGTGGATTAC CCACAGCGAT	15720
	GACCAAAGTA TGCGACACTT TCAAAAGCGT ATAACAATTG TTGATGGTCA AATTTCTAAT	15780

55

	CATTCCGATT ATCATTTTCAT ATAAAGAAGG TTTACATATT ATTAAAGATT TAATTGTTGC	15900
	GACATTACGA GCAGTTGTGC AATTAATCAT TTTGGGATTT TTGCTGCATT ATATTTTAA	15960
5	AATAACGAT AAATGGCTGC TTATTTTATG TGTATTGGTC ATTATTATTA ATGCATCATG	16020
	GAATACAATT AGTCGAGCAT CACCAAGTAT GCATCATGTG TTTTGGATAT CATTTCTAGC	16080
	TATCTTCATT GGAACGGCAT TACCGCTTGC AGGTACTATT GCGACAGGGG CCATTCAATT	16140
10	TACCGCAAAT GAAGTTATAC CTATCGGCGG CATGCTTGCA AATAATGGCT TGATTGCAAT	16200
	TAATTTAGCT TACCAGAATT TAGATCGTGC ATTCGTACAA GATGGTACTA ATATTGAATC	16260
	TAAATTATCA CTTGCAGCTA CACCTAAATT GGCTTCTAAA GGTGCAATAC GTGAAAGTAT	16320
15	TCGTTTAGCT ATAGTGCCAA CTATTGATTC GGTAAAAACA TATGGGCTTG TGTGATTCC	16380
	TGGTATGATG ACAGGCTTAA TTATTGGTGG CGTACCACCT TTACAAGCGA TTAAATTTCA	16440
20	ATTGTTAGTC GTGTTTATTC ATACAACCTGC GACCATTATG TCTGCTTTGA TTGCGACATA	16500
	TTTAAGCTAT GGTCAATTTT TCAATGCAAG ACATCAATTA GTAGCACGAA ATACTGATGT	16560
	TAAGAGTGAA TCATGATAGA TTTTACTGCA TCAGATTTAG GCATTAGTTT TAATTGGA	16620
25	TGAAGTGACG CGCACATATA GTATCGCTAT TCATTAGCGC AGCGAAAATA TTCATAAAGG	16680
	CACGCATACT TTGTAGTCAG TTATCTGTTC TGACATATAA AGCGTGCGTG CTTTTTTGGA	16740
	GTTATTGTTG AAAGTGAAGT AATTATACAT AATTATTAAA TGACATACTT GTGTTAATTT	16800
30	TTCAAATACT GAAAAACAAT TTCaATAATT TTCCaATTAA GCACAGAAAA TTAAAGCAAA	16860
	ATATTATATA ATAGAACGGT TATATATaAA nATTngTgCA CACATTTTTT AATAAATCGT	16920
	TATTCTAAGG GAAATGAATA TCGGAAATTT TGTTTGAAAG GAGTTTTTAAA TTGTCAATCA	16980
35	TGCGACTATT TACATTCAAT TTAAGTATTT TTATCGTAGG AATGGTTGAA ATGATGGTTG	17040
	CAGGAATTAT GAAGTTGATG AGTCAGGACT TACATGTATC AGAAGCTGTC GTTGGTCAAT	17100
40	TAGTGACAAT GTACGCTTTA ACATTTGCGA TATGTGGACC TATTCTGGTT AAATTAACGA	17160
	ACCGTTTTTC ATCAAGGCCT GTATTATTAT GGACATTACT TATATTTATC ATTGGTAATG	17220
	GCATTATTGC TGTAGCGCCA AATTTTTCaA TATTAGTAGT TGGTAGAATT ATCTCATCTG	17280
45	CAGCAGCAGC ACTAATTATC GTAAAAGTAT TAGCTATTAC AGCGATGTTA TCAGCACCTA	17340
	AAAATCGTGG TAAATGATT GGACTTGTCT ATACAGGGTT TAGTGGTGCT AATGTTTTTG	17400
	GTGTACCAAT TGGAACGGTT ATCGGCGATT TAGTAGGTTG GCGCTATACA TTTCTATTCT	17460
50	TAATTATTGT GAGTATTATT GTTGGCTTCT TGATGATGAT CTATTTACCG AAGGATCAGG	17520
	AAATACAACG AGGCCCTGTG AATCATGAGA CACCATCTCA TGAAAATCAT GTTACTTCGA	17580

55

CAAACTCAGT GACATTCGTC TTTATAAATC CACTTATTTT ATCTAATGGT CATGATATGT 17700
 CATTCGTTTC ATTAGCACTT CTAGTAAATG GAATCGCTGG CGTTATTGGA ACATCATTAG 17760
 5 GTGGTATATT CTCCGATAAA ATTACAAGTA AGCGTTGGTT AATGATTTCT GTTTCTATTT 17820
 TTATCGTCAT GATGTTACTT ATGAATTTAA TCTTACCTGG TTCAGGTCTA TTGTTAGCAG 17880
 10 GACTATTTAT TTGGAATATC ATGCAATGGA GTACTAATCC AGCAGTGCAA AGCGGTGTGA 17940
 TTCAACATGT TGAAGGCGAC ACAAGCCAAG TAATGAGTTG GAACATGTCT AGTTTAAACG 18000
 CTGGTATTGG TGTGGAGGC ATTATTGGAG GCTTGGTCAT GACACATGTT TCTGTTCAAG 18060
 15 CTATCACATA TACGAGTGCC ATCATTGGCG CATTAGGATT AATCGTTGTT TTCACATTGA 18120
 AAAATAATCA TTATGCTAAA ACATTTAAAT CATCATAATT CTCATATGAm AAGCACGCCT 18180
 GCTATCAAAT TCAGGTGTGC TTTTITAGAT GCGATAACGT TATTGATATG TGCGATAATA 18240
 20 GCGACGTTCA TTATGATACA TCGGCCAAGG CATTTTACCG CTTTITAGCAA AATTAGCTAA 18300
 ATCATTTTGC ATTTGTGCGAC TTAAAAATTT AAGGTGaGCA GTTGTTGGaT ATgAT 18355

(2) INFORMATION FOR SEQ ID NO: 68:

25 . (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1192 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

35 CGCAAAGAAG TACAAAAAAT GTTTTACAA GAAGGTATTA AAACACCTCA ACCAATTATG 60
 ACTGCTTATA ATCATAGTGA AAACGgTGTT TAGTAGTTTA TAATACATGG AGGTCATATT 120
 TAA[~]TGGCGTC AAAATATGGA ATAAATGATA TAGTAGAAAT GAAAAACAA CATGCCGTGTG 180
 40 GAACAAACCG TTTTAAGATT ATTAGAATGG GTGCAGACAT AAGAATTAAA TGTGAAAATT 240
 GTCAAAGAAG TATTATGATT CCACGTCAA CGTTTGATAA AAAACTTAAA AAAATCATCG 300
 AATCTCATGA TGATACACAA AGATAGGAGA ATGATTAATG GCTTTAACAG CAGGTATCGT 360
 45 TGGATTGCCA AACGTTGGTA AATCAACATT ATTTAATGCA ATAACAAAAG CAGGTGCTTT 420
 AGCAGCGAAC TATCCATTCC CTACGATTGA TCCTAATGTA GGGATAGTAG AAGTGCCAGA 480
 TGCTAGATTA CTTAAATTAG AAGAAATGGT TCAACCTAAA AAGACATTGC CGACTACATT 540
 50 TGAATTTACA GATATCGCTG GTATTGTGAA AGGTGCTTCA AAGGGAGAAG GGTTAGGTAA 600
 TAAATTTCTTA TCACATATTA GAGAAGTAGA TGCGATTTGT CAGGTCGTTC GTGCATTTGA 660

55

TAATATGGAA TTAGTACTAG CGGACTTAGA ATCTGTTGAG AAACGTTTGC CTAGAATTGA 780
 AAAATTAGCA CGTCAAAAAG ATAAGACTGC TGAAATGGAA GTACGTATTT TAACAACTAT 840
 5 TAAAGAAGCT TTAGAAAATG GTAAACCCGC TCGTAGTATT GACTTTAATG AAGAAGATCA 900
 AAAATGGGTG AATCAAGCGC AATTACTGAC TTCTAAAAAA ATGCTTTATA TCGCTAATGT 960
 TGGTGAAGAT GAAATTGGTG ATGATGATAA TGATAAAGTA AAAGCGATTG GTGAATATGC 1020
 10 AGCGCAAGAA GACTCTGAAG TGATTGTTAT TAGTGCAAAA ATTGAAGAAG AAATTGCTAC 1080
 ATTAGATGAT GAAGATAAAG AAATGTTCTT AGAAGaTTTA GGTATCGaAG AACCAGGATT 1140
 15 AGATCgrTTA ATTAGGAmCA ctTATGAATT ATTAGGnTTA TCCACCATAA TT 1192

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7494 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AATATAGCTG CAATAGCATC TCGTTTCATT TGTATAATCA ATTCCGGTTT AAATATCAGT 60
 GTGAACGTAA GCACGACACA GATTAAAAAT AACACTGCCG GAATGAGTCG TTTCAATCGT 120
 30 CGCTtCCAAA ACTCTAGCAA ATCGATTTTT TGCGTCCGAT AATACTCACT TATCAACAAA 180
 CTTGTTATTA AATAACCTGA AATAACGAAG AATGTATCTA CTCCTAAAAA GCCCCCACTT 240
 AACCATTGTG CATTCAAGTG ATAAATAATG ATTCCTATAA CTGCGAATGC CCTCAATCCA 300
 35 TCTAATCCAG GTAAGTATCG CGGGGAATAC ATTTTTTCTA AACGTTTAAA GTCTTTTGTA 360
 TCCA~~t~~GTTAA TAAACGCCCC ATTTATTTTT CTCTATTTTG TAGTATATCA CAATATTTTT 420
 GAAA~~A~~TAAAA TATTGCACTG aTTTTcATTA ATTGATTTAA CCCTTAATTA AGATAGTTTT 480
 40 AAATTTTTTA TTAAGTAGAA AACAATTATT ACAGTTGATT TCATTACTGC AAACCACATA 540
 TAAATTTGTC GATTTTACTA CATAACATAG ATTATCATAG ATTCTTGAAT TTTTAGCAAA 600
 45 ATAACGTGTA TTTTCATTAT ATTTTACAA AAAAAGGTTT GTTTTATATT TTATGCATCT 660
 TACTGTAACA GAATCATTAA GATATGCTAT TCGAATATAC TTTTCAAAA TTTATATAAT 720
 GAATAAATTA ACATGTATTG AAAAAAAGC GAAATGCAGC CTATCCTCTA ATGTAAACCA 780
 50 AACGATATAT CTCGTCAGAC TTTATATTTA AACGCTATGT GTCAC~~T~~TTTA AAATGAATAT 840
 TACTAAGATT GTCATATCAA TTATTATTGC ATCGAATTAA TCTTTTAAAT TTCTGTAATA 900

55

	ACGGAAGTCA TTATTAGAAT AAAAATACTG TGCACTAATA AATTTATCAA TTGTTCTTAA	1020
	ATAAATACCA TCGATATTTT GTTCTTTACA TGTCATTATA ACTTTATCTA AAAGTTTTTT	1080
5	ACCTATTTTT AAATTCCTAT AACCTTTATC AACAAACATT TTTTAAAGTG CAGACATATT	1140
	ATTATCTAGT CTAATCAAAC CTATAGTACC AACAAATATT TGaTGATTGT TTATTGCAAG	1200
	CCAAAATgCC CTCCATTATT CAAATAGTTA TGTTTCGATGT TCTCCAAATC AGGTTGATCA	1260
10	TCTCTATCAA TTTTATATATa AATTCATTTT TTTGAATCGA TAAAATAAAC TCGATTAGCT	1320
	CTTCCTTATA AGACCTATTA TATTCaATTA TGTTTATAGC CATTTTTATC TCCTTTTTCA	1380
	TTTAATTTAA TTATAAAATG TGCGTTTAGT TTGTATCTAG TGTACTCAGT ACAGCCTCAA	1440
15	ATGAAGTTTC ATTCCACTTG GCACTTAATA AAGACAAGTA TTTTAGCAGT AATACAATAA	1500
	AGTCCAATAA ATTTCCCTAA CTTCaATATC CACTTTTTAA AAAATGTATT TTTAATTAAT	1560
20	AAAAAACTC TCCCCAATTT CTATGGGAAG AGCTATATAT TTAATGTCTA AACATTACTT	1620
	TTATTTATTA TGAAGGAATT AGAATCCCCA AGCACCTAAA CCTTGTGCTT TGTATGCTTT	1680
	AACAGCTGCG TTGATTTGTT GGTCAACAGT GTTTGTTGGA CCCCACCTG GCATAGTTTG	1740
25	GAATAAACCT GAAGCACCTG ATGGGTTGTA AGCATTTACT TGACCATTG ATTACGAGC	1800
	GATGATTGCA GCCCATGTAG AAGCTGAAAC ACCAGTACGT TGAGCCATGA TTTGAGCTGC	1860
	TGATGAACCA GTAGCACCTG CAGTATTACC ATTGCTTAAT CTCACTGAAC TTGAAGTAGT	1920
30	TGAAGTGCTG TAGTTATGGT AAGTTGGAGC TGAACAGCT TCAACGTtTG AGTTACTTGA	1980
	TTGTGCATTG TAGCTTACTG ATTGTACATT TGAACCTTGG TTGTATGAAG TAGTGTAGTC	2040
	TGCACCTGCA ACGTTTGAGA AACCAGCAGT TTGACCATTA GCTGCTTCAT AGCTCCATGA	2100
35	CCATGTAGTA CCATTTGAAG TGAAGTTATA TTGGAAACCA TCTTTTACAA AGTGGATGTC	2160
	ATATgCACCA TCTTTGATTG GAGCTGCATT TAATTGATCT TGGTGATTAT GCGCTAAGTC	2220
	AACTAAGTGT GCTTGATCAA CGTTTACTTC AGCAGCGTGT GCTTGATGTC CTGTACCTGC	2280
40	TGCGTAACCT GTTACACCTA ATGCCACTGC TAATGATGAT GCCATAATTG TCTTTTTCAT	2340
	AGTAAAAAAT CCTCCAGTAA TAATTGTnAG TTTATGTTTT TAGTAATTAT AtTTTGaATT	2400
45	TGAATGTCGT AGTgCAAGTT TAAATTGTCT TTTATTTCTT TCaACGGTAC TCACTATATC	2460
	ACAaAAAACC AGCCAGTAAA TTACACTTTC TTTACAAAAC ATTACAATAT CAAGTGTTAT	2520
	TTGtAATGTT GAAATATGGC TGTTTTATAC TGTAATGTGA AATATGTGCC CTTTAGAATC	2580
50	CAATCAACCC TTGAAATAGT CTTTAACACA TAAGATTTTT ACTATATTTA GCTCAACTAT	2640
	TACAGCTTTC GTAATATTAC AGATTGTATT TTTGTTACAT AGCTGTAATA TATCTGACAT	2700

55

	TACACATGTA TTGATTGCTA TTATTGTTGT ATATTCAAAG TTTTAAAACA CACATCTTTT	2820
	GTGAATTGTC TTATCTTTTA TTAGCGCAAA TAAACTGCAG CTCAATTATA TTGTTCAACT	2880
5	TCATTCTCGC AATTCCACAAT AACATTAAAT AATTTTGGT CTCATATTTT CAAAAACAT	2940
	ACTGTTATTA TCCCATGAAT TTA AAAATAT CATTAGTATA TAAACGAAAC ACTTTACGAT	3000
	AAATGATATC TGCAAGCCAA GCTGTTACAA ATGGTACAAC AAAGAACGCT ACTACAATTA	3060
10	GTAAGACACT CAACCAAGCA GAATCAACCT CCATAAATT AAATGCATTA ATCGGTCCTA	3120
	CCATTCTAT AAAACCAAAT CCAGCTGACT CTTTCGTTCC ATGAATACCT ACTAATGCTG	3180
	ATACCAAACC TGATACAATG GCTGTCGTTA ATATTGGTAA CATAAGAAAT GGATATTTCA	3240
15	CCATATTAGG TATCATCATT TTAACGCCTC CAAAGAAGAC GGATAACGGC ACCCCTAAAC	3300
	GATTCACTTT ACTTGTAACA ATTATCAATA CTGCTTCAGT CGCGGAGATA CCAATTGACG	3360
20	CTGATCCAGC TGCTAAACCT GTAATACCTA TCGCAAAGGC AATGGCCACA GTTGATAGTG	3420
	GCGAAATAAT AATAAGACTA AATACCATTG AAATCAAAAT ACTCATGACA ATCGGTTGTA	3480
	ATTCTGTAAA ACCATTAACC ATATTACCGA TGGCTGTTGT AATCATTTTC GTATACGGCA	3540
25	ATATTAAAC ACCAATTGCA CCTGAAATAC CGCCAACAAC TGTTGGGAAT ACAATCAATG	3600
	CCATACTACC TACGCGATGT TGAATAAGTA AAATGAATAA CACTGCAATC GCTGCTGTAA	3660
	TCATTGTATT AATTAAATCA CCAATACCCG TAATCATCCA AGCACCATT TTAAGTGC	3720
30	CTGCACCGCT TCCTACATAT GCTGCACTTG CCACAACAGC AATTGCTAAT GGCGATAGGT	3780
	CAAATTTTCAT GGCAACCAAT GCACCAATCA AAGCAGGTAC TGTA AATTGA ATTGCAACGA	3840
	CAACGCCTAA TAACGTTTTA AAAATCGGAT GATAATCCAT AAAGTATTTA AAAATTTCTC	3900
35	CAAGTATCGC ATTAGGAACT AAACCCGCAA CAATACCTAT GGCGACACCT GATAAACTC	3960
	TAAATATAAA ATCTTTGGGT GTAATTGTTT TAATTGATGT CATAATATCA TCCTTCCATT	4020
40	TATGTATATA CATCTGTATG CAAATAATAA AGAGCCTTAA GTTATAAGCT GCCACTAGCT	4080
	TAAATTCTAA GATGTGCATG CCGATGTTGT TATATTTAGG CTAGCACTAT CATCTATAAC	4140
	TCAAGACTAT GAAAAATAGT ATATCACAAA ATTCTGAATT TTTAGATAAA TAAATTGGCA	4200
45	ATTTTTCAAA CATATTGTTA CAATACACTT TTATTTTATC TTCATTTTAA AAATCCATTA	4260
	ATACAATAGA AGAAAGACAT TCAATGCTT ACCAAAAAGG TACATTATTT GTTAGGAGCG	4320
	TATCAGCACT TACATATCAT CAACACAATT GACAATATAA TAGAAGATAC TGATAATAAG	4380
50	TGTTAAAACA ACAGATGTTA GGTAGTGAAC AAATGATGGA AAGTAAATCC ATAGATCCAA	4440
	GAATCGTTAG AACCAACAA TTGCTTGTCG ATGCTTTTCT TAAATTTCT AGAGAAAAGA	4500

55

	TTTACGCTCA TTTCGCTGAT AAAGAAGACC TCCTAGACTA CACATTATCT GTAACCATTT	4620
	TAAAAGACTT GAATGATAAT TTGAGCATTT CTAATGTCAT TAATGAAAAG GTTCTGCGTA	4680
5	ATATTTTCAT TTCAATTGCG AGTTATATCA AAGATGCTGC AAAGTCTTGC GAATTAAATA	4740
	GTGAAGCATT TTGCAACAAA GCACATCAAC GTATTAATAA TGAATTAGAA GATATTTTTG	4800
10	CGATTATGTT AGAAAACAGC TATCCGGAGC ATCAACGAGA TATCATTGTA AATAGTGCGA	4860
	GTTTTTTAGC AGCTGGTATC TCAGGCTTAG CATTACATTG GTTTAACACG AGTCAAGAGA	4920
	CAGCCGATGT GTTTATCGAT CGCAACCTTC CATTTTTAAT TCATCATATA GCACATTTTT	4980
15	AATAAACTT GGTATTTAGT CATGCATCTT GAAATCACTA TGTGACTTAG GTTCATACTT	5040
	GTACACACAA TAAAATTTAA CGTATTACGA TTGATTAGCC GTGTCTAGGA CATAAATCAA	5100
	CGTCCTATAC TCTACAATGT CATATTAGCA GTCGTTAACT GAATGAAAAT AAGCTTGTCA	5160
20	TTAAAACATA TAGATTTTAG TGACAAGCAT TTTTGTMTT GCGTACTTAA ACAACACTTC	5220
	AGGCAATATG TTGTTTAGGC AACAAATGAT ATGTGCGTGT TTATTGGCAA ACGTACGACA	5280
	TAGTAGTATA GTATGTCTAA ACAACATATG TTGCATAGTT GATATGCGTT GTTTAAATAC	5340
25	TAAGATAGGA GGGATTGACG TGAGCGAGAC AGATGAACCT CAGGGGTTTG AACGCACGCA	5400
	TAATATATTA AATATTAATC AGAGTAGTCT GGGTGTAGTG ACATACATTA CAAATAAATT	5460
	AAAGTCGACG TTGAAGCAAC ACATAATAAT TGCTCGTGGT AAAAAGCGAA TCGACTATCG	5520
30	ACTGTCTGAT AACTTTTACA TACGTATTAT GATAATGTAG AAATCAAGAA AATCGACTGT	5580
	GAATATACCT ATGCTATGCC CATTGCAATT TTAATAAGAC ACACGATGTC ATTGACAAT	5640
35	GCTCATTTCT TTGCTCAGTT ACGTCATCCT GTCTTATAAA ACAACATTGC AGACATGTAT	5700
	ATCAAACGAC ACTTCAATAA CATCACTTTG CCATCGTAC TACTAGTAAA ATCGTGTCTC	5760
	AAATCCCTTA TTTTAATTCC AAAAATCTGC TGGTCAAAAG ACCGAGAAAC TAAAAACATT	5820
40	ACTTAATGTG TTGATAAATT ACCATATAAA AATAATCTCA AAATATATCA ACACTTGATT	5880
	CTAAGGAGGA TATGACAATA TGAAAATTTT AGATAGAATT AATGAACTTG CAAATAAAGA	5940
	AAAAGTACAA CCACTTACTG TAGCTGAAAA ACAAGAACAA CATGCATTGC GTCAAGAcTA	6000
45	CTTAAGcATG ATCCGAGGAC AAGTATTAAC AACATTTTCC ACAATAAAAG TGGTTGATCC	6060
	AATCGGTcAG GATGTCACAC CAGATAAAGT TTATGATCTT CGCCAACAAT ACGGTTATAT	6120
	TCaAAATTAA tATTTGCTCA CGAGGTATTG CACTTAAGGT GCCAACTGAC CTCATAAACA	6180
50	AAGCCCATAC TGATTGAAGA CACTAATGTG tCsacCATGG TGCACATTAC GCTTCATCTC	6240
	TGTATGGGCT TTTTATTTAT TCTTTTGAGA ATTTCATTTT AGCAGACCAA AAAATTAAAA	6300

55

TGAACGACTG TGCCACCCGC TTCITTCAC TATTCCACCA ACTGGTCAAC TTCTTCATTT 6420
 GTGTTCCACAC CTAGAGAAAT CATCACTTCA TTTGGTTCAG TATTAAGGCT TTGCTGACTT 6480
 5 ACATTTTGAA AATGCTTGTh TTCTATTAAA ATTACGGkTG tTTGACCTAT tTGAATGCCG 6540
 ACCATTTTAT CTAACATTTG TGGGTTTCTA TTTATTTTAA ATCCTAACGC TTTATAAAAC 6600
 TGTGCGCTCT TTTCTAAATC TTGCACATGC AAATTAAACC ACATTGATTG AATCATGATT 6660
 10 GCACCCCAT TACTACTTAT TATAGTTTGT GACTTTAAGC CAATCACTTA ATGATAATCT 6720
 TGTGGGATTT ATTTTCAGCCA TTAATTCAA GTCTACTTCA TAACCTTTTT CTCCAACCA 6780
 TTGCTTTTCT GCAACACCAC TAACAAATTC TCCTTCTATA ACAGTAGATT TACCTGTCAC 6840
 15 TTCACTAAAA ATTGTTGCTG CTTCACTTAA TGTAACCTCA TCGGAACCAA TCTCTATTGA 6900
 TTGATGCGTA AAGCTTTGTG GATGTGCAAA AATATACGAT GCAATTTTAG CTATATCAAT 6960
 20 AGAAGAAATC ATTGTGAATT TTATATTCGG ATTAATAAAT TCTGGTAATG TAATACGTTT 7020
 ATCTTCGACT TTAGCAATGC GTAAAAAATT ATCCATAAAG AATGATGGTT TGATAACTGT 7080
 TGCATTTATA TTAGATTCCA TTAATCTATT TTCTATTTTT GCTAGTACTT CAAAGTGTGG 7140
 25 GCCAGTTCGA TTTCGATTAA CCCCTCCCGC AGTACTATAC ACAATATGTT GAATATTTTC 7200
 TTGCTCAGCT ATTTCAATTA TCTTCATACC TTGTCTTAAT TCTTCGCTAA CATCATCTTT 7260
 AACGATTGGC TGAATACTGT ATAAGCCATA CTTACCTTTC ATCGCTGATT GCAAACCTAAC 7320
 30 ATTATCACTC AGATCACCTT CArCGATTGA TAAATGCGGA TGTCTATGT CTGAAAGTTT 7380
 ACGATTATnC TTATTTCTAG TTAATGCACT TACATACCAT CCATCCTCTA ACAACTGTTT 7440
 TACAACTGCA TTACCTTGCT TCCCTGTTGC GCCTATTACn AAAATATCTT TCAT 7494
 35

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11802 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

AATTTATTTT GCCGTCCAC CCCAACTTGC ATTGTCTGTA GAAATTGGGA ATCCAATTTT 60
 TCTTTGTTGG GGCCCCGCC CAACTCGCAT TGCTGTAGA ATTTCTTTTC GAAATTCTCT 120
 50 GTGTTGGGGC CCCTGACTAG AATTGAAAAA AGCTTATTAC AAGCGCATT TCGTTCAGTC 180
 AATTACTGCC AATATAACTT CGTAGATCAT AGAACATTGA TTTATTTCCC AGCCTATTCT 240

	AGCAAAGGTA ATAATGATAT TAATAATGTA CAAAAAATAT AAATCAAATC GACATCCTTA	360
	TAAAACATCA GAACCACTAA AAACAAAAAA GCACAAAATA AAATTAAATT TAAAATAAAC	420
5	GACCACTTTT CAAAAAATC TCtTTTCaTa TTTCCACCCC TAATTTTAAT AAGCATTATT	480
	TTATATTCTC TTTTAAGTTT ATTATTCAAA AGGAAAACAG AAATATCTTT CaATATTATT	540
10	ATAAACATTT CAACTACTTT TAAAAACCAA CAAAAAATA CTTATTTTAA GTAGATGAGC	600
	ATAAGTGAAC ATAGTTCTTT AGTTATAATA ATTAATTCAA CCAAAGTCG ATTTGTTTTT	660
	GCAATTGGTT TTCATTTCTT CTTAAAGATA TTTTCATTAA ATCTGTCAA TCAATAGACG	720
15	CTATATTTT CAACTTATCT CTATATTTAT TTTTAGTACG TCTTTCTAAA TTTCCCCATT	780
	CCTCTTCTTC GTGAGTTAAT AAATGAAGCA TTGCTCGTTC TTGTATATTT TCAATCATTT	840
	TTAAATTCGG TTTTAAAAATA TGCAAATCAT CAAAACAATC TTTCCAACAA TCAACCATAT	900
20	CTCGTTTAA TTCAATTTC ACACGCCATA GAAATGTTGA ATCAATTTCA ACATCTGCAT	960
	TATCTTTACG TTCTTGTTTT TATTATAAAT CCGAATAAAC CTATCACTAT TACGCACACC	1020
25	AAAATATTTT GTTTCGTGGT TTACATTACG TCCATAAAAT ATAGTTTTCT TTACCGACTT	1080
	ATCTGACAAT GCATAATAGT CATTTAAATC AAATTCAAAA TCAAAGCCA AATCTAATCT	1140
	CGTAAACTA ACATCGTCCA AATAACTGAT GATATTTTGT TTTAACCAAA GCACTTCATC	1200
30	ATGCGAAAGC TTATTAGGAT TAAATTCAAC GCGCATATAC GTCTATTCCA AAGAGTTGCT	1260
	TTTATTTTGT CATATTCAAT ATAACTTTT TCTTTAAGAG CTTTAGCTTT AAAGTTTGTT	1320
	TGTAAATAT CCCAAGCCG AATTCAGGA TTAGTACTCA TAAAATGTGA AAGTCTCTCT	1380
35	GCGTTAGACA TGCTAAGATT CCCAACAATC GTTATAGCGT CAAAAGACAA TTTTGGATA	1440
	GCTAGTGACA TCCTATGTCG ATTTAACCGG CTATTACCGG ATATTAGAGT ATCCAGTTTT	1500
	ACAAATGGAT GAAACGAAAT TCAAAACACT AAAAAATATG TTCCACTAAC AGCAAAAAAA	1560
40	TACCATTATG TTCCTACTAA AAAACyAAAA ATACTGGAGA ACAAATGTCA GGATATAACT	1620
	TAGGATACTA TGTAATAAAA ATTTACAATA AAAAAACAGG AAAACAAATT TCAAGTAAAA	1680
45	GmATACCCAT ACAAAGAGGA TAAAATAAAA AACCTCGAAC TGaAATGATG ATCTTTTCAG	1740
	CTCGAGGTTT AAATATTGGT GCCTTATTTA TATAGATTCG TTATATTATA TTCTCTATTT	1800
	TCATTAACmT AATCCTTAAA GAGTTTTAAA TTAATACCTG CTAGATGATT CAAAAATGTT	1860
50	TCATCAACTT TTAAATAATT CAATAATTTT TGTGGTGTCA GTAAATnTCT ATCAAAATAC	1920
	AACTTTAATA AACTATTCTT TTTGACAGGA CGTGACATTT CAATCACGTC GTCTAAAGAT	1980
55	AATACTTTCT CGCTTTAnAC AAAnACAAAA ACTTACCCGA TTAAATCAA GTAAGTTTTA	2040

	TATTTGATAA AAAATCAATA AGTAATTGTG CGCCTTCAAC TTGAATATCT TTTACAACTG	2160
	GCGCGTCGAT ATACATATCA TACTGACCAC CGCCTACTGC ACGATAATTA TTTACACAAA	2220
5	TTGTATATGT CTGCTTTAAA TCAACTGCGT GACCTTGAAT CATCATATTG CTCACACGTT	2280
	GTCCCTTTGG TCTTCCAACA TGAATGGTAT AACTTACGCC ACCATATATA TCATAATTAA	2340
10	AGTGTGTGG TTTGGGTTCA AGGAAGTCTG CGCTCACACT AACTTCATCA TTTTTCACGT	2400
	CAAAATATTC TGCTGATCGT TCAATGGCTT CTTTAAGTTT GGCACCACTT ACAGCTAAAA	2460
	CTTTAAATGT ATTTGGAAAT GGGTAATTGT TAATAACATC TCGCATCGTC ACGACTTGCT	2520
15	TGAAACCACT AGCAGAATCA AACAAAGCTG TACAGGCAAC ATCTGCGTCA CTTTTTTCTA	2580
	ATAAAGCGTA ATTCATAAAA TTTGTAAAAG GATGCGGTGC CACACGTGCC TCAAATGCAT	2640
	GATTAATCGT CATATCATAT GGCAATGTAG TAATTTTCGT ATCTAACCAG TCCTCTAACT	2700
20	GCTTTCGTAA ATGTTGGTCA TCTTCATCAA TAGTAAATGT GGAATCATCT ATAACAGGAA	2760
	GTAATTCACA TGATTCAACG GATAGATTTT CATATTCATC AGTACTCAAG ACTACTCTGC	2820
25	CTACAGTTGT ACCTCTCGTA CCAGGTTGAA TCACAGCCGT TTGCTTAAAC CTTTCAGCAA	2880
	TTTGTGATG TTGGTGACCC GTAATAAAGA TATCTATATC TTTAGAAAAC GCTTCTAACA	2940
	TGGCATATCC TTCATTTTCA CCCGTTAATA CTTGCGTCCG CGTACCACTT TCTAAATCCT	3000
30	TTTCAAATCC ACCATGGTAA CAAACCACAA TGATATCTGC ATGTCGCTTC ATTTTCAGGTA	3060
	AGTATTGTTG AAGTATTTC AAGCACTAT GAAACGTAT GnCnTGAATA TGCTCTGGTT	3120
	GTTCCCAATG GGGGAATAAAT TGTGTCGTTA AACCTATCAC ACCAACAGTT TGATCTCCAA	3180
35	CCTGAAAATA CTTACACACG TTATCAGTCA ATGTACTATC ATTTTCATAT ATATTAGCGC	3240
	ACAAAACCTGG ATAATTGAGT CTGCGTAAAG TGTCTTTTAA GTATGGTAAT CCATAATTAA	3300
40	ATTCTATGATT ACCAAGCGTA CCAAAGTCGA ATGCCATTCT ATTATAAAAA TCAACTAAAG	3360
	GCTGGCTACT GCCGCTATGC GCGATTAAGT AATTACAAAA TGGTGACCCT TGCAAAAAAT	3420
	CACCATTATC TATTTTAAAA CTTTGGTCAT ACTGCCTTCT GTsTTGTTCT ATAACATGAT	3480
45	TCGCTAGTAA CAATCCCATA GGTTGATATT GATTTCTACT CGTAAAATCT GTTGGGAAAA	3540
	TATAACCATG TACGTCACTC ACGACATAAA ATGCTATGTT TGACATCCTC ACTCACTCCT	3600
	TCAATCACAA ACATCTTTCT TATTTCTATT ATATATTTAT TTGAAGTCTG TTGTAATCAA	3660
50	GGTTTTGTCA CCGAGTTTAA AACGAATCTT TGAACCTTCC ATACTTTCAA GTACTTTAGC	3720
	ATTGACCTTA ATTGTGACAT TTCCGTTTTC ATCTGCTTTA ACTGTTGGCA AAGTACTGTA	3780
55	ACCTGGTGGG TTATAATCGT TATCTTTACT TGAAAATTGT CCGATTTGAC GTCCGCCTTC	3840

	TATTGTCATT TCAAATGGCT CATTTACAGA AACATTTTGC GGGATATCAA ATGTTACTTT	3960
	TTCGTTCTGA TTTGGTGGTG TATGATCATC TGGTGTGTTT GGCTGAGGAT CTGCGCCTTT	4020
5	TTTCGCTGCCA TAACTACCTG CTTTAAATGT TGTTGGATCA TACCATTAT AACCCTCGG	4080
	CGGTTGTGAC CATGGCTCTT TTTCAGGCTC AGTTGAACGC TCTGGTCGTT CAAAATCAAG	4140
10	CAACTTAGTC TTTGTATCTA ATGTTAGGCT ACTCGCCTTA AGTGATTTCC CATCATTATC	4200
	TTTAGACATC CAAGCCGTTA TATTATTTAA TAGCTTACCG TTGTCTTGTT CTTTAAAACC	4260
	ATCATATGTT TTCTTCTTTT CTCCATTATC TTCTCTTACA TATTTGGGCG AACTATCTTC	4320
15	CACAAGTGAT GAATCACCGA TAAATGCTGC TTTACCTTTT CCAACTTTAG AAATTGCTAC	4380
	ATAGGGGCTT TCTGCTTTAC CGCCCCCATT ATAAATACCT TGATCTACAG CATGTGACCA	4440
	TTTACTTTTC GCTGGCAATT GTTCTGGTGT ATACACAATA CCTTTTGCTT TCTCTGGATT	4500
20	AGTAATTGCT AATGTCGATC CGGCATGCAT AGAGACAGAT TTCACACCTT CAGTAATACC	4560
	GAAACTTTCT TTTGAAGAAA CAATATTGCT CGTATTTAAA TCACCTAGTG CATTATATCG	4620
25	AAAACGTACG CCAAAGTTTG TAGATAACCA ATCTGAACTT TTCACACCTT GCATTGCAGT	4680
	AGAACTTTTT TCTTCTGCAT TCATACCTTT CGACATATCT TCATATGCTC CACGTCGATA	4740
	ACCATTCAAT GCCTCCGATG AATCAATACG ATTTAAATTT CGGTCAGCAT TGTAAATGATC	4800
30	TGAAATAAAG ACAACATTGC CACCTTGTTT CACATATTTA ACAATTGCTG CCTGTTCTGA	4860
	TTCTTTGAAA GGAATGTTAG CCTCAGGAAT TACAAATATT TTGGAACCTT TCAAACCTGC	4920
	TTCTGTTATG TTCGAATGAC CATCAATAGC TTTAACGTCA TAACCTTGTT TTTGTATTGA	4980
35	ATCCGCATAA TCTGAAAATG CACCATCACT AACCCAATCT GCAGCACCAG CTGTTTGACC	5040
	ATGAGAACGA TCGAATAATA CCGTTCGCTG TTGCTTTGTA GGTTCGATT CATGCGTTAT	5100
	AGCTAAAGAT TGCGGTAAAG CACTTAATGA TACCGTTGCA ACAATTGCAG AGACAGTTAA	5160
40	TGACTTATAT ATTTTTTTCA TTTTGTGAGG CTCCTTTTAA AATAAATTTG TTCTTGAATT	5220
	ATAGGATAAA AATTCGTTGC ATATGAGCAA TTTAACGAAA AATTTACAAA ATCTTATCAA	5280
45	ACTCTTAAAG AAAGTTATTA AAATTCATTT TTATAAATA CTTTTTAACA TTTAAATGTG	5340
	GTACGCTATA AGTGTAATTT CATTGCATAC ATATTACACG ATTAAGAATG TGAAGGGGAC	5400
	AGTTATCAAA TGAAAAATTT TAAGTGTTTA TTTGTATTAA TGTTAGCAGT CATTGTTTTT	5460
50	GCAGCAGCAT GTGGAACTC AAGTTCTTTA GATAATCAAA AGAACGCTAG TAATGATTCG	5520
	GATTCTAAAT CAGGAGGATA CAAACCTAAA GAATTAACCG TTCAATTTGT ACCTTCGCAA	5580
55	AATGCTGGAA CATTAGAAGC TAAAGCAAAA CCATTAGAAA AATTACTATC TAAAGAATTA	5640

	TCTAAAAAAG TTGATGTTGG TTTCTTACCA CCAACGGCAT ACACATTAGC ACATGATCAA	5760
	AAAGCAGCTG ATTTATTATT ACAAGCACAA CGTTTCGGTG TAAAAGAAGA TGGTTCAGCA	5820
5	AGTAAAGAAC TTGTAGATAG TTATAAATCA GAAATTCTTG TAAAAAAGA CTCAAAAATT	5880
	AAAAGCTTGA AAGATTTAAA AGGTAAGAAA ATTGCCTTAC AAGATGTAAC ATCAACTGCT	5940
10	GGATATACAT TCCCACCTTG C GATGTTAAAA AACGAAGCAG GTATTAATGC AACTAAAGAT	6000
	ATGAAAATTG TGAATGTTAA AGGTCATGAC CAAGCAGTTA TCTCATTATT AAATGGAGAT	6060
	GTAGATGCTG CGGCTGTATT TAACGATGCA CGTAATACTG TGAAAAAAGA CCAACCAAAT	6120
15	GTATTTAAAG ACACACGAAT TTTAAAATTA ACACAAGCTA TTCCGAATGA CACAATTTCT	6180
	GTAAGACCAG ATATGGATAA AGATTTTCAA GAAAAATTGA AAAAAGCTTT TATAGACATT	6240
	GCTAAATCAA AAGAAGGTCA CAAAATTATT AGCGAAGTTT ATTCACATGA AGGATACACA	6300
20	GAAACGAAAG ATTCAAATTT CGACATTGTA AGAGAGTACG AAAAATTAGT TAAAGATATG	6360
	AAATAATCAT TATTTAACAA ATGAATCATT AGCGAATTTG GTATTAAAAG CTTTCGTTCA	6420
	ATAGATATAT TCTAGATTAA TATTGAAAAG CTAGGCGCTA AACTGAAACA GATATAGAAA	6480
25	GGTGTGCTG TACATTTGAA ACCATTTGTA CACAGAAACC CAATGTCTAT GATATTTGAG	6540
	TTTACCTTGG CTTTTCTTTA TTAAAGAAAG GTGTCAAACA TGAGTCAAAT CGAATTTAAA	6600
30	AACGTCAGTA AAGTCTATCC TAACGGTCAT GTAGGCTTGA AAAATATTAA CTTAAATATT	6660
	GAAAAAGGTG AATTTGCAGT TATTGTCGGA CTATCTGGTG CTGGGAAATC CACGTTATTA	6720
	AGATCTGTAA ATCGTTTGCA TGATATCACG TCAGGTGAAA TTTTCATCCA AGGTAAATCA	6780
35	ATCACTAAAG CCCATGGTAA AGCATTATTA GAAATGCGCC GAAATATAGG TATGATTTTC	6840
	CAACATTTTA ATTTAGTTAA ACGGTCAAGT GTATTACGAA ATGTACTAAG TGGACGTGTA	6900
	GGTTATCACC CTACTTGGA AATGGTATTA GGTTTATTCC CAAAAGAAGA CAAAATTAAG	6960
40	GCAATGGATG CACTAGAACG CGTCAATATC TTAGATAAAT ATAATCAACG CTCTGATGAA	7020
	TTATCAGGTG GCCAACAACA ACGTATATCT ATTGCACGTG CGCTATGCCA AGAATCTGAA	7080
45	ATTATTCTTG CAGATGAACC AGTTGCTTCA TTAGACCCAT TAACTACGAA ACAGGTTATG	7140
	GATGATTTAA GAAAAATCAA CCAAGAATTA GGCATCACAA TTTTAATTAA TTTACATTTT	7200
	GTGACTTGG CAAAAGAATA TGGCACACGC ATCATTGGTT TACGTGATGG TGAAGTTGTC	7260
50	TATGATGGTC CTGCATCTGA AGCAACAGAT GACGTATTTA GTGAAATATA TGGACGTACA	7320
	ATTAAAGAAG ATGAAAAGCT AGGAGTGAAC TAACATGCCT TTAGAAATAC CTACAAAGTA	7380
55	TGACTCCCTT TTAAAGAAAA AGGTTTCTTT AAAAACGAGT TTTACCTTCA TGTTAATCAT	7440

	AATACCTCAA ATAGGTGATC TATTCAAACA AATGATTCCA CCTGATTTCG AGTATTTACA	7560
	ACAAATTACA ACGCCAATGT TAGATACCAT TCGAATGGcT ATCGTAAGTA CAGTATTAGG	7620
5	TAGCATCGTT TCAATACCAA TTGCGTTATT ATGTGCTAGC AATATCGTTC ATCAAAAGTG	7680
	GATTTCAATA CCCTCGCGCT TTATTTTAAA TATAGTTCGT ACTATTCCAG ATTTGTTATT	7740
10	AGCAGCAATC TTTGTGGCTG TATTTGGAAT CGGTCAAATT CCAGGGATAT TAGCACTGTT	7800
	TATTTTAACT ATCTGTATTA TTGGAATAATT ATTATATGAA TCATTGGAAA CGATAGATCC	7860
	AGGTCCAATG GAAGCAATGA CGGCTGTTGG CGCTAATAAA ATAAAATGGA TTGTTTTCGG	7920
15	TGTTGTACCA CAAGCCATAT CGTCATTTAT GTCATACGTA TTATATGCAT TTGAAGTAAA	7980
	TATACGTGCT TCAGCTGTGC TTGGATTAGT CGGCGCTGGC GGTATTGGAT TGTTTTATGA	8040
	TCAAACACTT GGTTTATTTT AATATCCAAA AACAGCAACG ATTATTTTAT TTACTTTAGT	8100
20	TATCGTCGTC GTCATTGATT ACATCAGTAC GAAAGTGAGG GCACATCTCG CATGACACAG	8160
	GAAATAGCAA AATATAATGT TCACACAAAA GCACACAAAC GAAAATTGAT TAAAAGATGG	8220
	CTTATTGCAA TTGTCGTCTT AGCTATTATC ATCTGGGCAT TTGCAGGTGT ACCAAGTTTA	8280
25	GAACTTAAAA GTAAATCATT AGAAATCTTA AAATCCATAT TCAGCGGATT ATTCCATCCT	8340
	GATATCAGCT ATATCTATAT ACCAGATGGC GAAGACTTAT TACGTGGTTT ACTTGAAACC	8400
30	TTTGCGATAG CCGTTGTAGG TACTTTTCATC GCCGCAATTA TCTGTATTCC ATTAGCATTT	8460
	CTAGGTGCAA ATAATATGGT AAAGCTACGC CCAGTTTCAG GTGTTAGCAA ATTTATTTTA	8520
	AGTGTTATAC GTGTCTTCCC AGAAATTGTA ATGGCACTTA TATTTATCAA AGCTGTTGGC	8580
35	CCAGGTTTAT TTTCAGGTGT ATTAGCTTTA GGTATCCATT CCGTAGtATG CTGGGAAAC	8640
	TTTTAGCTGA AGATATTGAA GGTCTAGATT TCAGTGCTGT AGAATCAITTA AAGGCCAGTG	8700
	GTGCGAATAA GATTAAAACA CTCGTATTTG CAGTCATACC ACAAATTATG CCTGCCTTTC	8760
40	TATCACTCAT ACTTTATCGC TTTGAACTAA ACTTACGTTT AGCTTCTATA CTGGGGCTAA	8820
	TTGGGGCTGG TGGTATCGGG ACACCACTCA TATTTGCCAT TCAAACACGT TCTTGGGACC	8880
	GTGTAGGTAT TATATTAATC GGTTTAGTAC TAATGGTCGC AATTGTCGAT TTAATTTCCG	8940
45	GTTCAATCCG AAAACGTATT GTTTAACATT AAATCAGGAT ACTCCTAAAT AAGAAGTCCT	9000
	ACCGTCTTAC GTTTCTCTAT TATAATAAAA ACAGCAGTGA AGAAAATAT TGTTATAGTT	9060
50	AACTTCACTG CTGTTTTTAT AATATCTAAA TTTATTCTAT TTCAATTCCT TTAAATAACT	9120
	TTTACCGAAC TCTGGTAATG TTACGTTGAA ATTATCTGCT ATAGTTGCAC CGATAGAACT	9180
	GAATGTAGTA TCACTTTCTA GTGCATGACC ACCTTTAAAT TTCGGACTGT ACATAATTAC	9240

55

	TGTAATAATT ACTAAATCGT CTTCTTTTAA GTTGCTAAAC AGTTCTGGCA AGCGATCATC	9360
	GAAATCTTTA ATTGCTTG TG CATAACCTGG TTTATCACGA CGATGACCGT ATAATGCATC	9420
5	AAAGTCTACT AAGTTTAAGA AGCTAATACC TGTGaaATCT TTCTTAACAA TTTTCATCAA	9480
	TTGATCCATA CCGTCCATGT TACTCTTCGT ACGAACCGCT TCTGTTACAC CTTCAACCATC	9540
	ATAAATGTCA TTAATTTTAC CGATGGCAAT AACATCATAA CCACCGTCTT TCAAATGATC	9600
10	TAAGACAGTT TTACCAAAG GTTTTAACGC ATAGTCATGT CGATTAGATG TACGTGTAAA	9660
	GTTTCCTGGT TCACCAACAT ATGGACGTGC GATAATACGA CCAATTAAAT ATTTAGGGTC	9720
	TTTTGTCAAC TCACGAACCT TTTCACAAAT ATCATATAAC TCTTCTAATG GGATAATGTC	9780
15	TTCATGTGCA GCAATTTGCA ATACTGGGTC TGCAC TTGTA TAAACAATTA AGTCACCAGT	9840
	TTTCATTTGG TGCTCGCCCC ACTCATCGAT AATTTGCGTA CCCGATGCCG GTTTGTTAGC	9900
20	AACAACTTTA CGACCTGTCA TTTCTTCAAT TTGTTGAATT AACTCTTCAG GGAATCCATT	9960
	AGGGTATACT TTAAAAGGTT GCATAATATT TAATCCCATTA ATTTCCAGT GACCAGTCAT	10020
	TGTATCTTTA CCAACTGAAG CTTCACTCAA TTTAGTATAG TATGCTTCTG GTTGTTCAAC	10080
25	TGCATTTACT ACTGGTAATT TATCGATGTT CCCTAGACCT AACTTTTCAA GGTTTGGTAA	10140
	AGTTTGATCG AAACCTTCTA AGGTATGTCT TAAAGTATGT GAACCTTCAT CTTTAAATC	10200
	AGCTGCGTCT GCGCCTTCAC CAATACCTAC TGAATCCATT ACGATTAAAT GTACACGATT	10260
30	AAATGGTCTT GTCATAGCTA TCACTCCCAA AATTTATATA TATTAGTAAT CTGAATCTGC	10320
	TTCTAAACCT TGCATAATT GAACACCTGC GCTCGCACCA ATACGTGTCG CACCTGCTTC	10380
	AACCATTTTA TTGAAATCTT CTAAATTACG TACGCCACCT GATGCTTTTA CTTCTACATC	10440
35	AGCACCTACT GTATCTTTCA TTAATTTAAC GTCTTCTGCA GTCGCACCGC CACCTGCAAA	10500
	ACCTGTTGAA GTTTTAACGA AGTCCGCACC AGCCGCTTTT GTTAATTCAC TCGCTTTTAC	10560
40	AATTTGTCGA TGGTCCAACA ATACCGTCTC AATAATCACT TTTACTGTGT GACCTTTCGC	10620
	AGCTTTAACC ACTGCTTCAA TGTCTTGTTG TACATCATCA AAACGTCCAT CTTTTAATGC	10680
	GCCGATGTTG ATGACCATGT CAATTTTCATC TGCACCATT TGAATTGCAT CTTCTGTTTC	10740
45	AAATGCTTTC GTTGCAAGTG TCGACGCACC TAATGGGAAT CCTATTACCG TACAAACGAG	10800
	CACCTCTGAA TCAGCTAGTC GCTCTGCTGC ATATTTAACA TGTGTTGGAT TCACACATAC	10860
	AGATTTAAAA TTGTATGctT TCGCTTCATC GATGATTGA TCGATTTGCG TACGTGTTGA	10920
50	CTCAGGCTTC AATAAAGTGT GATCTATATA TTTCTCAAAT TTCATACTTA CTACTCCTCG	10980
	TGTTATATAA TCTCTTTATT TAATTTTACT ATAAATACGA ATATATCTCG CGAATTTATA	11040

55

5 ATACTCATTAA AACCTAAAAT AATTAAAATA ATACCGAAAT GTGAACTTAA TGCATCATTG 11160
 CCTGGGAAAT TTAATGCTTT AAAATCGATT AGAGCCGCAG CAATCGCAAT ACCTACAGAT 11220
 10 ACCGCCACAT TAATAATTAA ATTATAAAAA CCAATAGCCA CACCTGTCAT ATTAAGATCT 11280
 ATTGTTTTTAA TGGCTTCGTT AAGTAAAGGT GCATACATTA AAGCAAAGCT ACCTGCAAAG 11340
 AATATCATAG AAATGACGAA GATTGAAATG TGATTACCTA CTGCAAATGC AGGTAAAATC 11400
 AAGCTCAGTG CTATTAAAAT AATTGCTGTG ATAATCGCTT GTTTTGAATT CAGATATTCTG 11460
 CCGATTTTAC CACTTAGTGC ACCAACAATG ACTGCTACTA TATAACCCGG TACTAATAAC 11520
 15 AGTGATGTTG TGTCTAGTTG CAGATGATAA ATTTGCTCCA TTATGAATGG GAACGTAAAA 11580
 ATATAACCCA ATTGGATAGC ATACATTACA AATACTATAA ATAAAAATGA AGCATAACGT 11640
 TTATTTTGGG AAAATGATTT ATTTACTAAT GGACGTTGCG CATTTTTAAT ATATAGCGCA 11700
 20 AAAACGATAA TCGCAATTAA GGCACCAATC ATATATAACC AATTAAAGTT CGTAATAAAC 11760
 AGCATGACTG TTGTAGCAGG GGATCCTCTA GAGTCGAnCC TG 11802

(2) INFORMATION FOR SEQ ID NO: 71:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1196 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

35 CTAAAGAAGA TGCGAAACAA GATGTTGATA AACAAGTTCA AGCTTTAATT GACGAAATCG 60
 ATCAAAATCC AAATCTAACA GATAAGGAAA AACAAGCACT TAAAGATCGT ATTAATCAAA 120
 TACTTCAACA AGGTCATAAC GACATTAACA ATGCGATGAC AAAAGAAGCA ATTGAACAAG 180
 40 CAAAAGAACG TTTAGCGCAA gCATTGCAAG ACATCAAAGA TTTAGTGAAA GCTAAAGAAG 240
 ATGCGAAAAA TGATATTGAT AAACGTGTAC AAGCTTTAAT TGACGAAATC GATCAAAATC 300
 CAAATCTAAC AGATAAGGAA AAACAAGCAC TTAAAGATCG AATTAATCAA ATACTTCAAC 360
 45 AAGGTCATAA CGACATTAAC AATGCGCTGA CTAAAGAAGA AATTGAGCAG GCAAAAGCAC 420
 AACTTGACAA AGCATTGCAA GACATCAAAG ATTTAGTGAA AGCTAAAGAA GATGCGAAAA 480
 50 ATGCAATAAA AGCCTTAGCT AATGCGAagc GTGATCAAAT CAATTCAAAT CCAGATTTAA 540
 CACCTGAGCA AAAAGCAAAA GCGCTCAAAG AAATTGACGA AGCTGAAAAA CGAGCACTAC 600
 AAAACGTTGA GAATGCTCAA ACTATAGATC AATTAAATCG AGGATTAAAC TTAGGTTTAG 660

55

TTGAAGCAAC ACCTGAGCAA ATCCTAGTTA ATGGTGAAC TATTGTACAT CGTGATGACA 780
 TCATTACAGA ACAAGATATT CTTGCACACA TAAACTTAAT TGATCAGCTT TCAGCAGAAG 840
 5 TCATCGATAC ACCATCAACT GCAACGATTT CTGATAGCTT AACAGCAAAA GTTGAAGTTA 900
 CATTGCTTGA TGGATCAAAA GTGATTGTTA ATGTTCTGT AAAAGTTGTA GAAAAAGAAT 960
 TGTCAGTAGT CAAACAACAG GCAATTGAaT CAATCGAAAA TCGGCACAA CAAAAGATTA 1020
 10 ATGAAATCAA TAATAGTGTG ACATTAACAC TGGAACAAAA AGAAGCTGCA ATTGCGnAAG 1080
 TTAATAAGCT TAAACAACAA GCAATTGGAT CATGTTnAAC AATGGCACCT GGATGTTCCA 1140
 15 TTCAGTTGAA GGAAATTTCA ACAACAAGGA ACAAGCGCCn GATTGGAACA ATTTGA 1196

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1519 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

CAATCGTTTC AACGCTATTA TCTTTAGACA ACAATTGTAA GCGTGTATGT GCAGTTTCTA 60
 AACAGTCTAT AATTGAGTT CTTAATTCAG CTGGATCATC TTTAAAAATA AAATCCATCG 120
 30 CTGCAACTTT GTAGACAAAT GTTAAATAGG TAAGTTCACT GTGACTCGTA ACGAAAATAA 180
 TGTTACCAAC TGGGTCATGC TTACGAATTT CACTGCCTAA TTTGATACCA TTAATATCAG 240
 TTGAAAGTTG AATATCTAAA AAGTAACAGC CTATGTCATT CATATTTTTA GCTTGCTCAA 300
 35 GCACCTCATA AGGATTATCA GTTGCGAGGG CAATTTCCAT AGGCTTTTCT TCTATCATT 360
 TATAATTTTT AATAATGGTA ACCATGTTTT CTCTTGTTT TGGATCGTCT TCGCAAATGA 420
 40 AAATTTTCAT ACATTCACAT CCTTATGGCT AGTTGTTAAT AATTTCAACT TTTGAATAA 480
 AGAAACCATT TCGATAATT GTATCTAATA AGACATTGTC TGCATTATCA GCAATTTCTT 540
 TTAAAGTTGA TAGACCTAAA CCACGACCTT CACCTTTAGT AGAAAACTT TCTTGAACA 600
 45 ATTCATGAAT GCGTGGTATA TCATCAGCGC ATTTATTCAT AACAATAAAC GTTACTGAAT 660
 TTTCACTTTC AATAAATGCA ACGCGAATGA TAGGGTCATC AATTTCAATT GATGCCTCAA 720
 TTGCATTATC AAGAATAATA CCAATACTGC GACTTAAATC GATCATATTC AAGTTAATGC 780
 50 TACTTACTTC ATCGGGTATT TCGATACTAA TCGGAATATT CATTCTTGT GCACGTAAAA 840
 TTTTCGCAGT AATTAAGCCT TTAATTTTAC GTACTTTAAG ATTCTCGATA CCATTTAATT 900

GTAGGCCAGG CATGTCATCT TCTCGAATGT ATTCTGAAAG TGTCGTTAAG ATATTGACAT 1020
 AATCATGACG GAACTTGCGC ATTTTCGTTG TGATAGCTTC AATCTTCAAT GTATATTCAT 1080
 5 AATAGGTTTC AATTTCTTCT TGATTACGTT TATATTTTCA CTCTTTAAGG AGAAATTGAG 1140
 AAATAACAAA TGTTAATATA CTTAAAAATA TAGTGATACC AATAAAAAATA AAAGAATACT 1200
 GCCTTATTAC TTTAGCTTCA TCCGAGTTTA TTTGTGAATA AAAGAAAAAT AATGAAAAAG 1260
 10 TAAGCAGTAA GATAGTCGAA ATAACTATTA AAAATCCTTT GTTTAGTATT AGATATGGTG 1320
 TGCTAATTTT TTTGAGAACT CTATTTATTA TATATGAGAA TAGTATACTA ATAGTCACAT 1380
 AAATAACAAA AAAGCTAGGG AATATTACAA ATATACTATC AGAAATTTTG GTGGATATAT 1440
 15 GCATATATAA CTATATACCT GTAGTTAGCA CnGtnATAGG AATAATCnGG CGAGGTCCAT 1500
 AATCCACCAA AATAGAATA 1519

20 (2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5445 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:
 30 GTAGGAATCT CTTTGTCTTT TTGGGAGGAC ATTTAATATG AATGTATATT TAGCAGAATT 60
 CCTAGGAACT GCAATCTTAA TCCTTTTTTG TGGTGGCGTT TGTGCCAATG TCAATTTAAA 120
 GAGAAGTGCT GCGAATGGTG CTGATTGGAT TGTCATCACA GCTGGATGGG GATTAGCGGT 180
 35 TACAATGGGT GTGTTTGCTG TCGGTCAATT CTCAGGTGCA CATTTAAACC CAGCGGTGTC 240
 TTTAGCTCTT GCATTAGACG GAAGTTTTGA TTGGTCATTA GTTCCTGGTT ATATTGTTGC 300
 TCAAATGTTA GGTGCAATTG TCGGAGCAAC AATTGTATGG TTAATGTACT TGCCACATTG 360
 40 GAAAGCGACA GAAGAAGCTG GCGCGAAATT AGGTGTTTTT TCTACAGCAC CGGCTATTAA 420
 GAATTACTTT GCCAACTTTT TAAGTGAGAT TATCGGAACA ATGGCATTAA CTTTAGGTAT 480
 45 TTTATTTATC GGTGTAAACA AAATTGCCGA TGGTTTAAAT CCTTTAATTG TCGGAGCATT 540
 AATTGTTGCA ATCGGATTAA GTTTAGGCGG TGCTACTGGT TATGCAATCA ACCCAGCACG 600
 TGATTTAGGT CCGAGAATTG CACATGCGAT TTTACCAATA GCTGGTAAAG GTGGTTCAAA 660
 50 TTGGTCATAT GCAATCGTTC CTATCTTAGG ACCAATTGCC GGTGGTTTAT TAGGTGCAGT 720
 GGTATACGCT GTATTTTATA AACATACATT TAATATTGGT TGTGCAATTG CrATTGTTGT 780

55

	CGAATCAATT TACTAAAATA AAAAGAAACG TAAATAGCAT AATTTAACAT GTTTGATTCA	900
	TGGATTATGC TATTTTTTCG CCAAAATTTA ACAGATTTTG TACAATGGGT TAGCGATTAT	960
5	TTTTTAATAA AGGAGATACT ACTAATGGAA AAATATATTT TATCTATAGA CCAAGGAACA	1020
	ACAAGCTCAA GAGCGATTTT ATTCAATCAA AAAGGGGAAA TTGCAGGGGT AGCACAACGT	1080
	GAGTTTAAGC AATATTTTCC ACAATCAGGT TGGGTGAAC ATGATGCAAA TGAAATTTGG	1140
10	ACATCTGTGT TAGCTGTAAT GACGGAAGTA ATTAATGAAA ATGATGTTAG AGCTGATCAA	1200
	ATTGCAGGTA TCGGTATTAC AAACCAACGT GAAACAACGG TTGTTTGGGA CAAaCATACT	1260
	GGCCGCCCAA TTTATCACGC AATTGTTTGG CAATCACGTC AAACACAATC AATTGTTCa	1320
15	GAATTAAAAC AACAAGGATA TGAACAAACA TTTAGAGATA AGACAGGATT ACTTTTAGAT	1380
	CCGTATTTTG CAGGTACAAA AGTTAAATGG ATTCTAGACA ATGTTGAAGG TGCACGAGAA	1440
20	AAAGCAGAAA ATGGCGATCT ATTATTTGGA ACGATTGATA CTTGGTTAGT ATGGAAATTA	1500
	TCaGgAAAag CtGCGCATAT TACTGATTAT TCaAATGCGA GTCGTACATT AATGTTTAAT	1560
	ATCCATGATT TAGAATGGGA CGATGAGTTA TTAGAActaT TACAGTACCT AAAAATATGT	1620
25	TGCCAGAAGT TAAAGCTTCG AGTGAAGTAT ATGGTAAGAC AATTGATTAC CACTTCTATG	1680
	GTCAAGAAGT ACCAATCGCT GGAGTAGCTG GTGATCAACA AGCAGCATTa TTTGGACAAG	1740
	CTTGCTTCGA ACGTGGTGAC GTGAAAAACA CATATGGAAC TGGTGGCTTC ATGTTAATGA	1800
30	ATACAGGTGA CAAAGCGGTT AAATCTGAAA GTGGTTTATT AACACAATT GCTTATGGTA	1860
	TTGATGGAAA AGTAAATTAT GCGCTGAAg GTTCCATCTT TGTTTCGGGT TCAGCAATCC	1920
	AATGGTTACG TGATGGATTA AGAATGATTA ATTCAGCACC ACAATCAGAA AGTTATGCGA	1980
35	CACGAGTTGA CTCTACTGAG GGTGTTTATG TTGTTCCAGC TTTTGTAGGT TTAGGAACAC	2040
	CATaTTGGGA TTCTGAAGCA CGTGGTGCGA TTTTCGGTTT AACACGTGGA ACTGAAAAAG	2100
40	AGCACTTTAT CCGTGCAACT TTAGAATCAC TATGTTACCA AACTCGTGAC GTTATGGAAG	2160
	CAATGTCAAA AGACTCTGGT ATTGATGTCC AAAGTTTACG TGTCGATGGT GGTGCAGTTA	2220
	AAAATAACTT TATTATGCAG TTCCAAGCAG ACATTGTTAA TACTTCTGTT GAAAGACCTG	2280
45	AAATTCAAGA AACTACAGCT TTAGGTGCTG CATTTTTGGC AGGTTTAGCA GTTGGATTCT	2340
	GGGAGAGTAA AGATGATATC GCTAAAAACT GGAAATTAGA AGAAAAATTC GATCCGAAAA	2400
	TGGATGAAGG CGAAAGAGAA AAATTATATA GAGGTTGGAA AAAAGCTGTT GAAGCAACAC	2460
50	AAGTTTTTAA AACAGAATAA ACTTGTAGAT TAGACTTTTG TATAAACATT GTGATACAAT	2520
	CAATTTAAGT TAATATTTGA ATCGAGAAGC GAGAGATTTG TTCGAACATG TACAATTGAA	2580

55

	GCATTGTCTA CTTTAAAGAG AGAACATATT AAAAAAGAATT TAAGAAATGA TGAATATGAT	2700
	TTAGTAATTA TTGGTGGCGG TATTACAGGT GCAGGTATTG CACTAGACGC GAGTGAAAGA	2760
5	GGAATGAAAG TTGCATTAGT TGAATGCAA GACTTTGCAC AAGGAACAAG CTCAAGATCT	2820
	ACAAAATTAG TCCATGGTGG TTTGCGTTAC TTAAAACAAT TCCAAATTGG AGTAGTTGCC	2880
	GAACTGGTA AAGAACGTGC GATTGTTTAT GAAAATGGGC CTCATGTTAC GACTCCAGAG	2940
10	TGGATGCTTT TACCAATGCA TAAAGGTGGA ACATTGGTA AATTCTCAAC ATCAATTGGT	3000
	TTAGGAATGT ATGATCGTTT AGCAGGTGTT AAGAAGTCTG AACGTAAAAA AATGTTATCT	3060
	AAAAAAGAAA CTTTAGCTAA AGAACCATTA GTTAAAAAAG AAGGTCTAAA AGGCGGCGGT	3120
15	TACTATGTTG AATATCGTAC TGACGATGCG CGTTTAACTA TTGAAGTTAT GAAGCGTGCT	3180
	GCTGAAAAAG GCGCAGAAAT TATCAACTAT ACTAAATCTG AACACTTCAC TTATGATAAA	3240
20	AATCAACAAG TAAATGGTGT TAAAGTTATA GATAAATTAA CTAATGAAAA TTATACAATT	3300
	AAGGCTAAAA AAGTGGTTAA TGCAGCAGGT CCATGGGTTG ATGATGTTAG AAGTGGTGAT	3360
	TATGCACGCA ATAATAAAAA ATTACGTTTA ACTAAAGGTG TACATGTTGT TATTGATCAA	3420
25	TCAAAATTCC CATTAGGTCA AGCAGTATAC TTTGATACTG AAAAAGATGG AAGAATGATT	3480
	TTTGCAATTC CACGTGAAGG AAAAGCGTAT GTAGGTACTA CAGATACATT CTATGACAAT	3540
	ATCAAATCTT CACCATTAAAC TACACAAGAA GACAGAGACT ATTTAATCGA TGCGATTAAT	3600
30	TACATGTTCC CTAGTGTTAA TGTTACAGAT GAAGATATTG AATCAACATG GGCAGGAATT	3660
	AGACCATTAA TTTACGAAGA AGGCAAAGAC CCTTCTGAAA TCTCTCGTAA GGATGAAATT	3720
	TGGGAAGGTA AATCAGGTTT ATTAACATTG GCAGGTGTA AATTAACAGG CTATCGTCAC	3780
35	ATGGCTCAAG ACATTGTTGA TTTAGTATCT AAACGCTTGA AAAAAGACTA CGGTTTAACA	3840
	TTTAGTCCAT GTAATACAAA AGGTCTGGCA ATTTACAGGTG GCGATGTAGG TGGTAGCAAG	3900
40	AACTTTGATG CGTTTGTTAG GCAAAAAGTA GATGTAGCTA AAGGATTTCG CATTGATGAA	3960
	GATGTTGCAA GACGTTTAGC ATCTAAATAT GGTTCAAATG TTGATGAATT GTTCAACATT	4020
	GCGCAAACAT CTCAATACCA TGATAGCAAG TTACCATTAG AAATTTATGT AGAACTTGTT	4080
45	TATAGTATTC AACAAGAAAT GGTATACAAA CCTAACGATT TCTTAGTTCG TCGTTCTGGT	4140
	AAAATGTATT TCAATATTAA AGATGTATTA GATTATAAAG ATGCTGTCAT CGATATTATG	4200
	GCAGATATGC TTGATTACTC TCCAGCTCAA ATTGAAGCAT AACTGAAGA AGTTGAGCAA	4260
50	GCAATTAAAG AAGCGCAACA TGGaAATAAT CAACCAGCAG TTAAAGAATA AtTAATTTGT	4320
	ACAATCATAA ACTGGTGTCC TGTTTTAAGG GCATCAGTTT TTTTATACGA GATACATTAG	4380

55

GTTATTAAAG GTGTGAGATG ATGACTGAAA AACAAATTAA ATTAAGTGA CAAGATAATA 4500
 CGAATATTGA AGTTAAAGTG AATTTTACAG ATGTAGATTC AAAAGGAATT ATTCATATAT 4560
 5 TTCATGGTAT GGCTGAACAT ATGGAACGTT ACGATAAATT AGCACATGCA CTTTCAAAGC 4620
 ATGGCTTCGA TGTGATACGT CATAATCATC GAGGACATGG TATTAATATT GATGAATCAA 4680
 CAAGAGGGCA TTACGATGAT ATGAAACGAG TTATCGGTGA TGCCTTTGAA GTAGCGCAAA 4740
 10 CAGTGAGAGG CAATGTTGAT AAACCATACA TTATAATCGG ACATTCAATG GGATCCGTTA 4800
 TAGCTAGATT GTTTGTAGAA ACATATCCGC AATATGTTGA TGGTCTAATT TTAAGTGGA 4860
 CTGGTATGTA TTCATTATGG AAAGGTTTAC CAACCGTTAA AGTGTTACAA CTGATTACAA 4920
 15 AAATTTATGG TGCTGAGAAA CGAGTTGAAT GGGTTAACCA GTTAGTATCA AATAGTTTAA 4980
 ATAAAAAAT ACGTCCATTA CGTACACAAA GTGATTGGAT TTCTAGTAAT CCAATTGAAG 5040
 20 TAGATAaCTT TATTAAAGAT CCATATAGTG GaTTTAAATGT GTCAAATCAA TTATTATATC 5100
 AAACAGCCTA TTATATGCTA CATAATCAC AATTAAAAAA TATGAAAATG TTAAaTCATG 5160
 CCATGCCTAT ATTATTAGTT TCAGGATATG ACGATCCTTT AGGTGATTAT GGTAAAGGGA 5220
 25 TTTTAAATTT GGCGAATATA TATAGAAACG CTGGCATnAA AAATGTTAAA GTGAATCTTT 5280
 ATCATCATAA ACGTCATGAA GTGTTATTTG AAAAnGATCA TGACnAAATT TGGGAAGACT 5340
 TGTTTAAATG GTTGAATCAA TTTTATAAAA AATAAAGAAA GTGGAATTAA ATATGAATAA 5400
 30 AAATAAGCCT TTTATTGTAG TAATTGTGGG GCCAACTGCT TGCAG 5445

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
 35 (A) LENGTH: 2569 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TGGCTTGAAC TACGCCAATA AGTCCCCCTA GTACAAGAAT GAATACCATG ATATCGACCG 60
 45 CTTCTATCGT ACCTCAACC ATGCTACTTG TTATTTGTTT TGGTCCAGCT GGATGTTGCT 120
 TTAATCTTTC ATAAGTATTC GGAATTGATA CCGGCTTATT AATTGCACCT GATTTAAATT 180
 GTTCAATCTT AATTTTAACC CCCATTTTGT CTAGTTCCTG TTGCGTACCC GGAACCTTTT 240
 50 TCACTTGGTT ATGAGGGTTA ACTATCTTTA GTTCTTGGGA TGAAGGTTTC TAAGAAAGTT 300
 TAGAATATGC ACCAGCAGGA ATAACCCATG TTGCTATAAC TGCAACAACC GTTAAATGA 360

55

	TAATTGTATT TTCCACGGTT TCATCTCCTT CGACATTAA CCTAGCATTT CTACCTTAAA	480
	GATTTTATAA ATATAAATTA AGAAAGTGCA CCCCGCATCA AAATAGAGGC ATTATTTTCA	540
5	GGGGGTGCAC ATAAATAATA AAAATCATGC ATTTGACATA TAGTAATTGA AAAGCGTTTC	600
	AATTCAATTA CTTTTTAATC ACAGTACCTA CTTTACCCTC TAAGGCAGCA TCTAATTCAT	660
	TTAATGATGT TATAAGCACA CTTCTTTTGT GATTGTTTTT AATAAATGAT ATGGCTGCTT	720
10	CAATTTTGG TAACATACTT CTTTTTGCAA ATTGATTTTC GTCTATATAT CGTTTAAATT	780
	CATCAACATT TGTGTGTTTC AAAGGCTGTT GGTTTTCAGT GTTAAAATTA ATATATACAT	840
15	AATCAATTGC TGTTAAAATA ATCAATTGAT CGCATGGAAT ATTAGCACCC AACAACGCAC	900
	TTGTTTTATC TTTGTCTATA ACTGCATCAA TACCTTAAA ACCATCATGT TGCTCTCTAA	960
	TTACTGGTAT ACCTCCACCA CCAGCAGCAA TAACGAGTGT ATCATTTTTA ATAAGTGTTT	1020
20	TAATACTCTC TAATTCAATA ATAGAGATGG GTTGTGGTGA AGGAACAACG CGTCTATATC	1080
	CTCTTCCAGC ATCTTCAACA AATATAAATC CTTTTTCTTT TTGAATTTGT TCAGCTTCTT	1140
	CTTTGTGTGA AAATAACCCA ATTGGTTTTG AAGGATTGTT AAATGCCGGA TCATTTTCAT	1200
25	CAACTTCAAC TTGTGTCACT AGTGTTACCA CTTGTTTATC CATTCOAATA GAATGCAATT	1260
	CATTTTGTA GCTTTCTTGT AATTGATAGC CGATGTAAGC TTGACTCATT GCGCCACATT	1320
	CAGCAAATGG AAATGCCGGA CCTTGGTTAT GTTCTGCAGC ATAGTTAAGT CCCAAATTAA	1380
30	TGCTTCCAAC CTGTGGTCCA TTACCATGAC TAATAACAAT CTCATGTCCT TTTGTnATTA	1440
	AyCCTACTAA TGATTtCGCA GTATTTTTAA CAAGCTCGAG TtGgTyCTTG aGGTGATTn	1500
35	CCTAAAGCAT TACCACCTAA TGCTACTACT ATTTTCGCCA TCATATTCAC TTCCTTATAT	1560
	CATTTAAAAT TCACCCAATG TAGCAACCAT GaCTGCTTTG ATTGTATGCA TTCTGTTCTC	1620
	AGCTTCTTGG AATACAACTG AAGCTTTACT TTCGAATACT TCATCTGTAA CTTCATTTC	1680
40	TCGAATACCA TATTTTCAA AAATTGTTG ACCTATTTTC GTATCAGCAT TATGGAAAGA	1740
	TGGTAAGCAA TGCTCAAAAA TAACATTTGG ATTACCAGTT TTATCCATTA TTTCTTTATT	1800
	TACTTGATAT GGTTCOAATA ATTCAAGTCG TTCTTTCCAT ACTTCATCAG GTTCACCCAT	1860
45	TGATACCCAA ACATCAGTGT AAATTACATC CGAACCTTTT ACaCCTTGGT CaATATCATC	1920
	TGTGATTAAT ATGTTGCCaC CATTTTCaGC GGCAATATTT TTACAGCGAT TTAATAATTC	1980
	ATCTGTTGGA TTTAATTCTT TTGGACAAAC TAAATGGAAG TTCATACCCA TAATGGCAGC	2040
50	ACCTTGCAAT AATGCATTTG CAACGTTATT ACGACCATCT CCAACATATG TAAAGTTAAT	2100
	ATCTGCATAA TCTTTTTTTA AGACTTCTTT TGCTGTAAAG AAATCAGCAA GAACTTGAGT	2160

55

TTCTACTGTT CTTTGTGAAA AACCACGGTA TTCAATGCCA TCATACATTC CACCAAGCAC 2280
 ACGTGCAGTA TCTTTAGTTG TTTCTTTTTT ACCCATTTGT GATCCAGTTG GGCCTAAATA 2340
 5 AGTTACATTT GCACCTTGAT CATGCGCTGC AACTTCAAAT GCACATCGCG TTCTTGTA GA 2400
 ATCTTTTTCA AATAACAGTG CAATATTTTT ATTTTAAAC ATAGGCTTTT CAGTGCCAAT 2460
 ATATTTAGCA CGTTTTAAAT CCTCGGAGAG TGTTAATAAG GTTCTACCTC TTGTCGTGAA 2520
 10 AAGTCTAATA AAGTTAAAAA ACTTCTGTTT CGTAAATTTT TCATTAAaA 2569

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1273 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CCTGGAACCA TCCaATCGtG CaAATCtTGa AAGaGAATAC GCAACAACAA TTAAATGTAT 60
 25 TGGAACACTA TATTCCAAAT GACCATCCAG CACTCGTTGA ATTAAAAATA TGGGAACGTT 120
 GGTACATAA ACAAGGTTAC AAAGACATCC ATTTAGATAT TACTGCGCAC CACCTAGATC 180
 CTATTACACA GGTtATTTA TTCAATGTCA TTTGCTGAA AATGAATCTC GAGTTTAAAC 240
 30 AGGTGGTTAT TACAAAGGAA GCATCGAAGG GTTTGGATTA GGATTAACAC TTTAAGTAAG 300
 GGAGTATGCA CAATGTTAAG AATCGCCATA GCCAAAGGAC GTCTAATGGA TAGTTTAATT 360
 AACTATTTAG ATGTAATTGA ATATACGACA TTATCAGAAA CATTAAAAAA TAGAGAACGC 420
 35 CAATTATTAT TAAGTGTAGA TAATATTGAA TGCATTTTAG TAAAAGGAAG TGACGTGCCA 480
 ATCTATGTGG AACAAAGGAAT GGCAGACATA GGCATTGTTG GTAGCGACAT ATTAGATGAG 540
 40 CGCCAATATA ATGTTAATAA TTTGTTGAAT ATGCCTTTTG GAGCATGTCA TTTTGCGGTT 600
 GCAGCGAAAC CTGAAACGAC CAATTATCGT AAAATCGCAA CGAGTTATGT TCATACTGCT 660
 GAAACATATT TTAAATCAAA AGGTATTGAT GTCGAATTGA TTAAATTGAA TGGCTCTGTT 720
 45 GAATTGGCCT GTGTTGTAGA TATGGTAGAC GGAATTGTCG ACATCGTTCA AACAGGTACT 780
 ACGCTAAAAG CGAACGGACT GGTGAAAAG CAACATATTA GTGATATCAA TGCAAGATTA 840
 ATAACATAA AAGCAGCTTA TTTTAAAAAA TCACAATTAA TAGAGCAATT TATTCGCTCT 900
 50 TTGGAGGTGT CTATTGCCAA TGCTTAATGC ACAACAATTT TTAAATCAAT TTTCATTAGA 960
 AGCACCATTA GATGAGTCAT TGTATCCaT TATTCGCGAT ATTTGTCAGG AAGTTAAAGT 1020

TTTAGaAATT AGTCATGAmC AAATTAAAGC AGCATTTGAC ACATTAGATG AAAAAACAAA 1140
 ACAAGCATTa CAACAAAGTT ATGAAAGAAT TAnAGCATAT CAaGAAaGTA TtaAACAGaC 1200
 5 GaATCAACAG TTAGAAGaAT CAGTGGaGTG tTrTGaAATA TACCATCCmC taGaAAAGTG 1260
 CCGTATTtIAT GTG 1273

(2) INFORMATION FOR SEQ ID NO: 76:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1308 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

20 GTTGATAAAT TAAAAATGTT TTTATCAGAT ATTCAAAGTT ACCAACAATA TAGTAAAGAT 60
 CATCCGGTGT ATCAGTTAAT TGATAAATTT TATAATGATC ATTATGTTAT TCAATACTTT 120
 AGTGGACTTA TTGGTGGACG TGGACGACGT GCAAATCTTT ATGGTTTATT TAATAAAGCT 180
 25 ATCGAGTTTG AGAATTCAAG TTTTAGAGGT TTATATCAAT TTATTCGTTT TATCGATGAA 240
 TTGATTGAAA GAGGCAAAGA TTTTGGTGAG GAAATGTAG TTGGTCCAAA CGATAATGTC 300
 GTTAGAATGA TGACAATTCA TAGTAGTAAA GGTCTAGAGT TTCCATTTGT CATTTATTCT 360
 30 GGATTGTCAA AAGATTTTAA TAAACGTGAT TTGAAACAAC CAGTTATTTT AAATCAGCAA 420
 TTTGGTCTCG GAATGGATTA TTTTGATGTG GATAAAGAAA TGGCATTTC ATCTTTAGCT 480
 TCGGTTGCAT ATAGAGCTGT TGCCGAaAAA GAACTTGTGT CAGAAGAAAT GCGATTAGTC 540
 35 TATGTAGCAT TAACAAGAGC GAAAGAACAA CTTTATTTAA TTGGTAGAGT GAAAAATGAT 600
 AAATCATTAC TAGAACTAGA GCAATTGTCT ATTTCTGGTG AGCACATTGC TGTCAATGAA 660
 40 CGATTAACTT CACCAAATCC GTTCCATCTT ATTTATAGTA TTTTATCTAA ACATCAATCT 720
 GCGTCAATTC CAGATGATTT AAAATTTGAA AAAGATATAG CACAAATTGA AGATAGTAGT 780
 CGTCCGAATG TAAATATTTT AATTGTGTAC TTTGAAGATG TGTCTACAGA AACCATTTTA 840
 45 GATAATGATG AATATCGTTC GGTTAATCAA TTAGAAACTA TGCAAAATGG TAATGAAGAT 900
 GTTAAAGCAC AAATTAAACA CCAACTTGAT TATCGATATC CATATGTAAA TGATACTAAA 960
 AAGCCCTCAA AACAACTCTGT TTCTGAATTG AAAAGACAAT ATGAAACAGA AGAAAGTGGC 1020
 50 ACAAGTTACG AACGAGTAAG GCAATATCGT ATCGGTTTTT CAACGTATGA ACGACCTAAA 1080
 TTTCTAAGTG AACAAGGTAA ACGAAAAGCG AATGAAATTG GTACGTTAAT GCATACAGTG 1140

GATGGATTAA TCGATAAACA TATTATCGAA GCAGATGCGA AAAAAGATAT CCGTATGGAT 1260

GAAATAATGA CATTTATCAA TAGTGATTAT ATTCGATATT GCTGAAGC 1308

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1431 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

GATGCCATTn ATnnGTATGC AAGAAGTTGT TCCGGGTTCA GGTGGATTaC CAGTTGGTAC 60
 TGGTGGTAAG ACGTTACTAA TGCTTTCAGG CCGTATAGAC TCACCAGTTG CTGGGATGGA 120
 AGTGATGAGA CGTGGCGTAA CAATTGAAGC GATTCATTTT CATAGTCCAC CATTTACAAG 180
 TGATCAAGCA AAAGAAAAAG TTATTGAATT GACACGTATT TTAGCTGAAC GTGTTGGACC 240
 AATTAAATTG CATATTGTAC CATTTACAGA ATTGCAAAAA CAGGTAAATA AAGTTGTACA 300
 TCCAAGATAT ACAATGACTT CAACGAGACG TATGATGATG CGTGTTGCTG ATAAATTAGT 360
 ACATCAAATA GGGGCTTTAG CTATTGTAAA TGGTGAAAAC CTAGGGCAGG TAGCCAGTCA 420
 AACACTTCAT AGCATGTATG CAATTAATAA TGTAACCTCT ACTCCTGTAT TACGTCTTTT 480
 ATTAACCTTAC GATAAAGAAG AAATTATTAT TAAATCGAAA GAAATTGGTA CATTTGAAAC 540
 ATCTATTCAA CCATTTGAAG ATTGTTGTAC AATTTTCACC CCTAAAAATC CAGTAACCGA 600
 ACCAAACTTT GATAAGGTAG TCCAATATGA AAGTGTCTTT GATTTTGAAG AGATGATTAA 660
 TCGTGCTGTT GAAAATATTG AAACACTTGA AATAACTAGT GATTATAAAA CTATTAAAGA 720
 ACAGCAAACA AACCAATTAA TAAACGACTT TTTATAAATA AAATCCTAGA GTAAATTTAA 780
 ACATAAGGGG ATGTTAAACT ATGGATTTGA ACTTAACGAT GATTATAATC ATAATTTTAT 840
 TTGGTTTTAT CGCGGCGTTT ATAGATTCGG TTGTAGGGGG TGGCGGTTTA ATTTCTACGC 900
 CAGCATTATT AGCAATCGGT CTACCACCAT CTGTGGCTTT AGGTACAAAT AAATTGGCAA 960
 GTTCGTTTGG TTCTTTAACT AGTACGATAA AGTTTATAAG GTCCGGTAAA GTGGACTTAT 1020
 ATGTTGTTGC CAAATTATTT GGTTTTGTAT TTTTGGCATC TGCATGTGGC GCATATATTG 1080
 CAACGATGGT TCCGTCACAA ATATTGAAAC CTTTAATCAT CATTGCACTT TCGTCGGTGT 1140
 TTATATTCAC ATTACTTAAA AAAGATTGGG GCAATACACG CACGTTTACT CAATTTACAT 1200
 TTAAGAAAGC CATAATATTT GCAGCACTTT TTATATTAAT CGGCTTTTAT GATGGATTG 1260

TAAGTGCAGC AGGAAATGCT AAAGTTTGA ACTTTGCTTC TAATATAGGT GCGCTTGAT 1380

TATTTATGGT ATTAGGACAA GTAGATTATG TAATAGGTTT AATTATGGCT A 1431

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4403 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AATATTATTT TAAATTCAAT ATTTATTGGT GCATTTATTT TAAACTTATT ATTCGCCTTT 60

ACCATTATTT TCATGGAAAG ACGTCTGCC AATTCTATCT GGGCTTGGTT ACTAGTCTTA 120

GTTTTCTTGC CTTTATTCGG CTTCAATTTA TACTTACTAT TAGGACGACA AATTCAACGT 180

GACCAAATTT TCAAAATTGA TAAGGAAGAT AAAAAAGGAT TAGAGTTAAT CGTTGATGAG 240

CAATTAGCTG CTTTAAAAAA TGAAAACCTT TCAAATCCA ATTATCAAAT TGTAATAATTT 300

AAAGAAATGA TTCAAATGTT GTTATATAAT AACGCAGCAT TTTTAACAAC AGACAACGAT 360

TTArrrrtAT ACACAGACGG CCAAGAAAAA TTTGATGACC TAATACAAGA CATCCGTAAT 420

GCTACTGATT ATATTCATTT TCAGTACTAT ATTATTCAA ATGATGAATT AGGTCGTACC 480

ATTTTAAATG AACTTGGTAA AAAAGCGGAA CAAGGTGTAG AAGTTAAAT TCTTTATGAT 540

GACATGGGTT CTCGTGGACT GCGTAAAAA GGCTTACGCC CGTTTCGCAA TAAAGGTGGA 600

CATGCTGAAG CATTTTTCCC ATCAAAATTA CCTTTAATTA ACTTGCGTAT GAACAATCGA 660

AACCATCGAA AAATGTGTTG AATAGATGGG CAAATTGGAT ATGTTGGTGG TTTAATGTT 720

GGTGATGAGT ACTTAGGTAA ATCAAAAAA TTCGGCTATT GGCGAGATAC GCATTTACGA 780

ATTGTCGGGG ATGCAGTGAA TGCATTGCAA TTACGATTTA TTCTAGATTG GAATTCACAA 840

GCCACACGTG ACCACATCTC CTATGATGAT CGTTATTTCC CAGATGTAA TTCTGGTGGA 900

ACAATTGGCG TTCAAATAGC TTCTAGTGGT CCTGACGAAG AATGGGAACA GATTAAATAC 960

GGCTATTTGA AAATGATTTT ATCTGCTAAA AAATCGATTT ATATTCAATC TCCCTATTTT 1020

ATACCTGATC AAGCCTTTTT AGATTCTATT AAAATTGCGG CATTAGGTGG TGTGATGTC 1080

AATATCATGA TTCCTAATAA ACCTGACCAT CCGTTTGT TTGGGCTAC TTTAAAAAT 1140

GCAGCATCCT TATTAGATGC CGGTGTAA GTATTTCACT ACGACAATGG CTTTTTACAC 1200

TCAAAAACAC TTGTTATAGA TGATGAAATT GCAAGTGTGG GAACAGCTAA TATGGACCAT 1260

AAATTAAAC AAGCTTTTAT AGATGATTTA GCAGTATCTT CTGAATTAAC AAAAGCACGT 1380
 TATGCTAAGC GAAGTCTTTG GATTAAATTT AAAGAAGGTA TTTCACAATT ATTGTACCTT 1440
 5 ATCTTATAAA ATAGAAATAT GAGGAGTGTA aCTTTAATGC AACAATCAGA CGTCATTAGT 1500
 GCTGCCAAAA AATATATGGA ATCTATTCAT CAAAATGATT ATACAGGCCA TGATATTGCG 1560
 CATGTATATC GTGTCACTGC TTTAGCTAAA TCAATCGCTG AAAATGAAGG TGTTAATGAT 1620
 10 ACTTTAGTCA TTGAACTCGC ATGTTTGCTT CATGATACCG TTGACGAAAA AGTTGTAGAT 1680
 GCTAACAAAC AATATGTTGA ATTGAAGTCA TTTTATCTT CTTTATCACT ATCAACCGAA 1740
 GATCAAGAGC ACATTTTATT TATTATTAAT AATATGAGCT ATCGCAATGG CAAAAATGAT 1800
 15 CATGTCACTT TATCTTTAGA AGGTCAAATT GTCAGGGATG CAGATCGTCT TGATGCTATA 1860
 GGCGCTATAG GTGTTGCACG AACATTTCAA TTTGCAGGAC ACTTTGGTGA ACCTATGTGG 1920
 20 ACAGAACATA TGTCACTAGA TAAGATTAAT GATGATTTAG TTGAACAGTT GCCACCATCT 1980
 GCAATTAAAC ATTTCTTTGA AAAATTACTT AAGTTAGAAT CTTTAATGCA TACAGATACG 2040
 GCGAAGATGA TTGCTAAAGA ACGTCACGAC TTTATGATGA TGTACTTGAA ACAGTTTTTTT 2100
 25 ACGGAATGGA ATTGTCACGA CTAGACATTG AAGTTGTAGT ATGATGATGC GATGTAATGG 2160
 CGTGTGTGTG TGGAAAGCTTG GTGTCATGCC ATGTTACTTT GATGTGTTGT TGTGGGAGCT 2220
 TGGTGACATG TCATGCTACT TTGATGTGCT GGTACCACGA TGCGTCTTGA TGTAGTGCTA 2280
 30 TGATGTGGCA TTGCGGTGTT ATGGTGTAT AGACAGGTTT GCGGTTGATG CCATGTTACT 2340
 TTGATGTGCT GGTACCACGA TGCGACTTGA TGTAGTGCTA TGATGTGGCA TTGCGGTGTT 2400
 ATGGTGTAT AGACCGGTTT GATGTTGATG CCATGTTACT TTGATGTGCT GGTGCTACGA 2460
 35 TGCGACTTGA TGTAGTGCTA TGATGTGGCG TTGCGCTGTT ATGGTGTAT AGCCAGGTTT 2520
 GGTGTTGATG TCATGCCGTT ACGATTCTAT GATATGTTGT TGGGACGTTG CAATGTGTAT 2580
 40 TATGCCGTTG TGACGTTATT ATTTACACT GTTACATGTA TAAGTGAATT GCTGTGGAAA 2640
 TTTGCGACAT AACTGCTAC ACTGATGAAT CATTGTGTCA AGATGACATT GCGATGAAGA 2700
 ATGACAACCTC TGTTATTAAC CACTTTTTAC AACTGAAAA CTCGTTAATA TTATTTCAAA 2760
 45 TAAAAACAGC AGTAGGATGA CTTTCACATT TGAAATCATC TTAGTGCTGT TTCTATTTAT 2820
 CACATATTGT ATAATGTGAC ACTAAGTTTC GCTATTGAAG CGAAAAATAA TGTGCGCCCT 2880
 ATAAAGTTAA AATTATCTTC AACTTTTAGG GTGCACATTA TTTGGACTTG CTAAGGTTAT 2940
 50 TTCTTTTTCT TTTAGACAC AACTTGTGTG TTTTGCCTT TTTTATTGct GCCGCCGTTG 3000
 TGCTCTCTTT CATACGCTTC AATGAAAGGT TGTACTTCTT TTTTAGCGAC TTTTTCATAA 3060

55

CCAAGTGCTG ATGCTGAGCT TAATGAAATC CAGATAATCA TAATTGGTGA AATGACCATC 3180
 ATCATGTAAC CCATTGACG TTGTTCTGCT GGCATCGTTT TACTTGATAC ATATGCTTGG 3240
 5 ATAAAGTATA AAACACCGGC AATAATTGTA ATCCAAATAT CAGGACGTCC TAAATCGAAC 3300
 CATAAGAAGT GTGGATATTT AAACAAACCA TCTACAAGTT GGTCTTTAAG TACAAAGTAT 3360
 AATCCCATGA TGATTGGTAA TTGGATTAGC ATTGGTAAAC AACCCAACAT ACTCTTAATC 3420
 10 GGGTTCATGT CATACTTTTT ATATACTTGC ATTAATTCTT GGTTCGAGC CATTTTTTCT 3480
 TCTTGTGTAC GCGnCaCGTT cACTTTTTCT TGAATTTTTT CAACTTCTGG CTTTGCAACT 3540
 TTCATTTTTT GACGCATCAT ATGACTATTT TTATAGTTTG ACAACATGAA TGGTAATAAA 3600
 15 ATAATACGAA TTACCAATAC AAGGATAATA ATAGCTAAAC CATAATTGTC GTTTAATAAG 3660
 TTATTTCCCA ACCAATCCAA TACATTTTTT ATTGGATCTA CGAATGTATT GTAGAAAAAy 3720
 20 cwCtACGTTT TTCAGGTTTA GAATAGTCAC AACCAGCCAA AAAGACCATA ATACCTAAAA 3780
 ATAATGGTAG TAACGCTTTT TTCTTCATTT TTCCACCTCT ATCATTATAT TCACATAGGA 3840
 TTTATTCTAT CACATTAATG AGTACGTATG AAACAATAAG TGGAAAAATT TAACTAATTA 3900
 25 TTAAAAAAT CTTTGAATCG ATTAACAGTC TTTTCAATAT TTTCACTTT AGAAATGGCT 3960
 GAAATGACTG AAATTCCATT GGCACCTGCT TCTACAATCG GCGCCACATT ATTAGTATTG 4020
 ATACCGCCAA TAGCTACAAT CGGTAGTTGC GGATTCATTT CTTTAAACGT TGCAATCATT 4080
 30 TCTGGACCTA CTGGTATATG CGCGTCATGC TTCGACGGCG TAGGATAGAT TGGTCCAACA 4140
 CCTATATAAT CmACATGAGT TAAATCAGAT TTTGCATACT CATCTAAATC ACTAATACTA 4200
 AGTCCAATAA TTTTATCAGT GAAATATTGT GCTATCTCTT TGACTTTCGC ATCATCTTGA 4260
 35 CCGACATGTA TACCATCCGC GTTAATTTCT TTTGCCAAGG ATACATCATC ATTAACGATA 4320
 AAAGGCACAT CATATTGATG ACAGAGATGC TGTAATTCTT TAGCTAATAC AAGTTTATCG 4380
 40 TTTCCTTTTA AAGCTGATTC ACC 4403

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1808 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

TGGAnCCAAT ATTAGAAATG ATTAAACAT TAACAGGTAT TAATAGTCCT TCAGGAGnCA 60

55

	TAACAAATAA AGGTGCGTTA TTAATAACAG TGCCAGGCAA AAATGATGAA GTACAACGCT	180
5	GTATTACTGC TCATGTTGAT ACTTTAGGTG CaATGGTTAA AGAAATTAAA GAAGATGGTC	240
	GCTTaGCAAT AGAATTAATT GGAGGATTCA CGTATAACGC GATTGAGGGT GAATATTGCC	300
	AAATTAAAAC TGATGCTGGT CAAATATATA CAGGAACAAT TTGTCTGCAT GAAACAAGTG	360
10	TTCATGTATA TAGAAATAAT CATGAAATAC CTAGAGATCA AAAGCATATG GAAATAAGAA	420
	TTGATGAAGT AACTACATCA GAAGAAGATA CAAAGAGTTT AGGTATTTCA GTAGGTGATT	480
	TTGTTAGCTT TGATCCACGT ACAGTTATCA CGTCATCAGG TTTTATTAAA TCTCGTCATT	540
15	TAGATGATAA AGCTAGCGTA CGgTtGATAC TACAATTACT AAAGAAATTA AAAGAAGAGC	600
	AAATAATATT ACCACATACA ACGCAATTTT ATATTTCTAA TAACGAAGAA ATAGGTTACG	660
	GTGCAAATGC ATCAATTGAT TCGAAAATCA AAGAATATAT TGCATTAGAT ATGGGCGCGT	720
20	TGGGAGACGG TCAAGCATCG GATGAATATA CAGTTTCTAT TTGTGCCAAA GATGCTTCAG	780
	GTCCATATCA TAAGCAATTG AAATCGCACC TAGTTAATCT TTGCAAATA AATAACATTC	840
	CATATAAAGT AGACATATAT CCATATTATG GTTCAGATGC TTCAGCAGCT TTACATGCTG	900
25	GTGCGGATAT CAGACATGGT TTATTTGGCG CTGGCATTGA ATCATCTCAT GCAATGGAAC	960
	GAACACATAT TGATTCTATT AAAGCGACAG AGAAATTACT ATATGCATAT TGCTTATCAC	1020
30	CAATTGAGTA AACCAATTAGT GTTGACAAAT GTGaACGACC TATGTAATAT AATGAACTAT	1080
	AAAAATAATT AGAATTTTCT AAAGAAATAG TAGCAGATAT GAAACGTAGC AAATAGAAAG	1140
	CTAATGGGTG ATGGGAATTA GCACGCCATA TCTTGTGAAT TGGACTTTGG AAAACAATTG	1200
35	AATGAGTTTT GAAAGTGAAC ATGAATTATG TTAAC TAAGG TGGCACCACG GTAACGCGTC	1260
	CTTACAGGTA TATGCGTTAT GTGGTGTCTT TTTATTTAGA CAAATGTAG TAGTTAATTA	1320
	AAGGTAGCAA CAGAAAGTTA GTGGATGATG TGAAC TAACA CCGAGATTAA TGAAATTGGG	1380
40	TTTTGTCTGC AACAGAAAAA TTATATATAG TAAAGAGTGA ACTATGAATA TTTCGAATAT	1440
	TCGGTTAATT TAGGTGGTAC CACGCGTCAC nTCCTTTATA TTGATAAGGA TGCTGGCGCT	1500
	TTTTTGAAAG GAGCGTATAG AATGGATATA TTTTATAAAA AAATAAAAGC AAATGTAACG	1560
45	CCCGAAGTTT TAGCACAAC TCAATCCAAG AAGaTCATTT TGGAAAGTAC AAATCAACAA	1620
	CAAACTAAAG GTCGCTATTC AGTTGTTATT TTTGATATTT ATGGCACTTT AACTTTAGAT	1680
50	AATGATGTAT TATCAGTAAG TACTTTAAAA GAATCGTATC AAATCACTGA AAGACCGTAC	1740
	CATTATTTAA CGACTAAaAT AAATGAAGAC TACCATAATA TTCCAAGATG AGGCAACTTA	1800
	AGTCATTA	1808

55

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1320 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

10	TGGTCGTCAA TTTCTTGATT ATATCTATAA TCCTCATTTT CAATATTAGA GTCTGTAGAA	60
	TCATCGATAT TATTATCATT CGCATGACTA GAAGCAGAAT CATTATTTT ATCATTGCTT	120
15	TCTTCTTTT TGAAGTCTT ATTTATCAAG TAAATTTCTT CATCAAAATC AGCTTGTTGA	180
	GATGTATCAT CTTTATTTTG ATTAGAAAAA TGTGTAGCCT TTGATCTTTT TCTTTGCCGT	240
	CTTTTCTTAG ATGTATTCCT CGTAAATAAT TCTAATTCAT CTTTATCTTC ATTTGATTCT	300
20	TGTTGATCGT TCTTCGTTTT ATCATCCATC AATACTCACA CCCTTTAATA AGATGGTAAA	360
	TGGGCACGGA ATCTTTCAAT AAATTTCTCT CCACGCTCTT CAAAAGTACT ATATTGATCC	420
	CAACTCGCAC AAGCAGGTGA CAATAATACA ACATCATTTG GTTCTATAAT ATCTTGTAAT	480
25	TTATCAACAG CGTCTTCGAC ATTGTTTCGCT TCAATGACCG ATTTCCCTTG ACTATTACCT	540
	AGTTTAGCAA ACTTAGCTTT CGTTTGTCCG AATACAACCA TCGCGCGAAC ATTTCCATA	600
	TAAGGAATGA GTTCGTCAAA TTCATTCCCT CGATCCAAAC CACCACATAA CCAAATGATT	660
30	GGTTGATTAA ATGAATTTAA GGCAAAGTGT GTTGCTAGCG TGTTTGTTGC TTTGGAATCA	720
	TTATAATATT TATTAGTTCT ATTAGTACCA ACATATTGCA ATCTATGCTC TATTCCTGAA	780
	AATGTAGTTA AACTATCAAT AATTGCTTTA ATAGGTACAC CAGCanaATA CAAGCAAGCA	840
35	CAGCTGCTAA TATATTTcTA AATTATGTTT ACCAGGCAAT ACTAGAtCTT CAGTGTTAAT	900
	AATaCGAACA CCTTTATaAA CGATAAAACC ATCTTtAATA TAAaTACCAT CArCTtCTTG	960
40	TTGAGTTGAG AAATACAATG TCTTAGCTTT TAATTCTTCC GACTCTATCA CTGTCTTTG	1020
	ATGATAATTA CAAATCAAAT AATCCTCTTC CGTTTGATTT TTATATATTT GCTTTTTAGC	1080
	ATTTTGATAG TTTTCTAAAT TTTCATGGTA ATCTAGATGC GCCGAATAAA TGTTAGTAAT	1140
45	TATAGCAATG TGTGGTTTAT ACTTTTCGAT TCCAAGTAAC TGGAATGACG ACAACTCTGT	1200
	AACTAAATAA TCTGTAGGCT TTAATTCTTG TGCTACTTTA GATGCAACAT AACCAATATT	1260
50	GCCGGATAAT CTTCCAGTTA AGCGACTTTT TTAAACATA TCTCCAATTA GAGAAGTAAC	1320

(2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4280 base pairs

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

	TTTACACCAA TCAAAAAATC GAACTGATAT AAATAAGTAC AAAGCTTATC TATCAATCCG	60
10	ATTTAGTTAT AAAACAAAAA AAGCCACAGT AATGTGGCTT TTTGTTATAT TCAGTATCAA	120
	AATGGTATCA ATAGCCATT TCGGAAGTCA AGAATGGCTT AACACGCGG TTAAAGCTA	180
	TCCAATACTA CCTTCCATT CGAAGTTGAT TAAACGGTTC ATTTGACCG CGTATTCCAT	240
15	TGGAAGTTCT TTTGTAAATG GTTCGATGAA TCCCATAACA ATCATTCTG TCGCTTCTTC	300
	TTCAGAAATA CCACGACTCA TTAGATAGAA TAATTGTTCT TCAGAACTT TTGAAACCTT	360
	GGCTTCATGT TCTAATGATA TTTGATCGTT GAATACTTCG TTATATGGAA TTGTATCTGA	420
20	TGTTGATTCG TTATCTAAGA TTAATGTATC ACATTCAATA TTTGAACGAG CACCTTTTGC	480
	TTTACGTCCA AAATGAACAA TACCGCGATA AATACTTTA CCACCATTTT TAGAAATAGA	540
	TTTAGAAACA ATTGTAGAAG ATGTATTAGG TGCTTTATGA ATCATTTTAG CACCGGCATC	600
25	TTGAACTTGT CCTTTACCAG CAAATGCAAT AGATAATGTA CTACCTTTTG CACCTTCACC	660
	TAAAAGAACA CAGTTTGGAT ATTTTCATCGT TAACTTAGAA CCTAAGTTAC CATCTACCCA	720
30	TTCCATATTT CCGTTTTCAT AAACAAAAGT ACGTTTTGTA ACTAAATTGT ATACATTGTT	780
	CGCCAGTTT TGAATCGTAG TATAACGAAC GTGCGCATCT TTATGCACAA TGATTTCCAC	840
	AACAGCAGAG TGTAAGAAC TAGTTGTATA AACTGGTGCA GTACAACCTT CTACGTAATG	900
35	TACAGAAGCA CCTTCATCAG CAATGATTAA TGTACGTTCA AATTGACCCA TGTTCCTCAGA	960
	GTTAATACGG AAATAAGCTT GTAGTGGCGT ATCTAGTTTG ATATTTTTAG GTACATAAAT	1020
	GAAGGAACCA CTGACCATA CTGCTGAGTT TAACGCCGCA AATTTGTTAT CTGCTGCAGG	1080
40	TACTACAGAA GCAAAGTATT TTTTGAATAA TTCTTCATTT TCTTGTAAG CACTATCTGT	1140
	ATCTTTAAAG ATAATACCTT TTTCTTCAAG TTCTTTTCC ATATTATGGT AAACAACCTC	1200
	AGATTCATAT TGAGCAGAAA CACCAGCTAA ATATTTTTGT TCAGCTTCAG GAATTCCTAA	1260
45	TTTATCGAAA GTTCTTTTAA TTTCTTCTGG CACTTCATCC CATGAACGTT CAGCTTGTTT	1320
	TGAAGGCTTT ACATAGTAAG TAATGTCATC GAAATTCAAT TCTGATAAGT CGCCACCCCA	1380
50	TTGAGGCATT GGCATTTTAT AAAACAATTT TAATGATTTA AGACGGAAAT CTAACATCCA	1440
	TTCCGGCTCA TTTTTCATGT TAGAAATTTT TCTAACGATA TTCTCAGTTA AACCACGTTT	1500
	TGATCTGAAA ATGGACACAT CATCGTCGTG GAATCCATAT TTATAATCCC CAACATCAGG	1560

55

	TTTAATTCAT	GATGTAAACC	ATATTATAAC	AATGACATGA	CATCTTATAA	AAATTTTAT	1680
	ACTTTTATAT	GTCTAATATC	AAAATTATCT	ATGATTAACA	GCATTCTATT	CTTCTTCAGT	1740
5	CGTACCTTCT	GCTTTACCTT	CTTTAGCAAC	AGTACCTTTT	TCCAATGCTT	TCCAAGCTAA	1800
	TGTGGCACAT	TTAATACGAG	CTGGGAATTG	AGATACACCT	TGCAATGCTT	CAATATCTCC	1860
	CATTTCTTCT	GTAATCACAT	AGTCTTCACC	AAGCATCATT	TTCGTAAATT	CTTGGCTCAT	1920
10	TTGCATTGCT	TCTCCAAGTG	AATGACCTTT	AACAGCTTGT	GTCATCATCG	ATGCACTTGC	1980
	CATTGAAATC	GAACAACCTT	CACCTTCAAA	CTTAGCATCT	TTTATAATGC	CGTCTTCTAT	2040
	ATCAAATGTT	AGTCGTATAC	GGTCACCGCA	TGTCGGGTTA	TTCATATCTA	CTGTCATAGA	2100
15	CCCGTTATCT	AATACACCTT	TATTTCTAGG	ATTTTTATAA	TGATCCATAA	TGACAGATCT	2160
	ATATAATTGA	TCTAGATTAT	TAAAAATTCAT	AAGAGAAAAA	CTCCTTCGTT	TGTTTCAAGG	2220
20	CATTTATTAA	CTGATCAACG	TCTTCTTTCTG	TGTTGTATAT	ATAAAACTC	GCTCTAGCTG	2280
	TTGAAGACAC	ATTTAACCAT	TTCATTAACG	GTTGCGCACA	ATGATGCCCA	GCTCTAACCG	2340
	CTACACCTTC	TGTATCTACG	GCTGTAGCAA	CATCGTGTGG	ATGTACATCT	TGTAAATTAA	2400
25	ACGTIATTAC	ACCTGCACGA	CGATCCTTTG	GCGGGCCATA	AATTTCAATT	CCTTCAATTG	2460
	CAGACATTTG	CTCATAAGCA	TATATCGTTA	ATTCTTGTTT	ATATTTATGA	ATTGCATCAA	2520
	AACCTATGCG	TTCTAAATAG	CGAATAGCTT	CTGCAAGCCC	AATTGCTTGA	GCAATTAATG	2580
30	GAGTACCCGC	CTCAAATTTA	GTAGGTAAAT	CAGCCCATGT	TGCATCATAC	TTACTTACAA	2640
	AATCAATCAT	GTCGCCACCG	AACTCAATCG	GTTCCATTTT	TTGTAGTAAC	TCACGTTTAC	2700
	CAAATAATAC	GCCAATACCT	GTTGGTCCAA	GCATTTTATG	ACCACTAAAA	CTATAAAAAAT	2760
35	CAGCATTTCAT	TTCTTGATA	TCAAGTTTCA	TATGTGGTGC	TGCTTGCGCC	CCATCAACAC	2820
	TGATAATTGC	ACCATGTTGA	TGAGCTATTT	CTGCAATGGT	TTTAACATCA	TTAATTGTAC	2880
40	CGAGCACATT	AGATATATGT	GCAATAGCAA	CGATCTTTGT	TTTATCATTA	ATCGTTTGCT	2940
	TAATATCCTC	GATGTTTAAAT	TCACCGTCAG	CTGTCAATTGG	TATAAATTTT	AATGTCGCAT	3000
	TTTTACGCTT	TGCTAACTGT	TGCCAAGGAA	CAATATTGGC	ATGATGTTCC	ATTTCAAGTGA	3060
45	CAACAATTTT	ATCGCCCTCT	TCAACATTTG	CATCACCATA	GCTATGTGCT	ACAAGGTAA	3120
	TCGACGCAGT	TGTTCCGCGT	GTAAAAATGA	TTTCTTCAAA	ATACTTCGCA	TTAATAAAAC	3180
	GACGAACGGT	TTCACGGGCA	TTTTCATAAC	CATCAGTTGC	CAATGATCCT	AATGTATGAA	3240
50	CACCACGATG	AACGTTTGAA	TTATAACGCT	TGTAGTAATC	TTCTAAAAA	TTTAACACTT	3300
	GCACAGGCGT	TTGACTTGTC	GCTGTTGAAT	CAAGATATGC	TAAACGTTTG	CCATTGACTT	3360

55

5
 10
 15
 20
 25

CTTCATTACAC	GACCTTTCTT	AAATAAAAAAT	CCTAATCATT	TAAATACTGA	CGTTGTATTA	3480
GTCTTATACC	AATATCGACA	GTCTATATCT	ATTACAAACT	TTTATTTTCA	AAATATTATT	3540
TAGAAACTTT	GCGTTCAATT	ACTTCTCTCA	ATTGACGTTT	AACGTCTTCG	ATAGGTAATT	3600
CACGTACTAC	TGGATCTAAG	AAACCATGTA	TAACAAGACG	TTCCGCTTCT	CTTTGAGAAA	3660
TACCACGACT	CATTAAATAG	TAAAGTTGAT	CTGGATCAAC	ACGACCTACT	GATGCAGCAT	3720
GACCAGCTTG	TACATCATCT	TCATCAATTA	ATAAAATAGG	ATTGCGCTCA	CCACGAGCAT	3780
G TTCAGATAA	CATTAATACA	CGTGATTCCT	GATTAGCAAT	TGATTTAGTT	CCACCATGCT	3840
TAATGTAGCC	GATACCATT	AATACAGACG	ATGCATGTTT	TTTCATAACA	CCATGTTTAA	3900
GGATATAACC	ATCTGTTTCT	TTACCATATT	GTACGATTTT	AGATGTTAGA	TTAATTTTTT	3960
G TTCGCCTGT	ACCTACAACT	ACTGATTTAA	GTGAACCTGT	TGAACGATCA	CCAAATAAAT	4020
TTGTTGTATT	ATCAATAATT	TGGCTACCCT	CATTCATTAA	ACCTAGTGCC	CAATTAATTG	4080
AGGCATCCGC	TTAGTAATA	CCACGTCGAA	TGATATGACC	TGTAAAGCCT	TTATCCATAT	4140
AGTCCACTGA	GCCATATGTG	ATATTTGAAT	TTGCACCAGC	AATCACTTCA	GAAATAATAT	4200
T LAATTGATT	TCCTTCACCA	GATGCATTTG	mTAAGTAATT	TTCAACATAT	GTGACTTCGG	4260
CGCTTTCTTC	AGTAACGATG					4280

(2) INFORMATION FOR SEQ ID NO: 82:

30
 35

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15598 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

40
 45
 50
 55

TCnGACTCGA	ACGGTGmAAC	TAttCCGTTG	TaATTCCgGA	GgAAaCAAGG	TATGCCCATC	60
TGCaAAGAAA	gaATGsAATG	AACTTTTTGG	AAATGTAGAA	GTGGTAAATA	AAGATAAAGG	120
ATATTACATT	CTGAGAAGTA	TAAAAGCTTG	AAATGAAATG	GATATTCTGT	TATAGTTATA	180
TAATGTAAAA	ATTTATGTTT	AATAAGTGTG	TACTTTTACG	TTAAATAGAT	AAGTTAATTA	240
AGAATAAATA	TAGAATCGAA	AATGGTGTCA	TCATTAGTGT	TGCCGTTTTT	TTTTTGTTCT	300
TTTATTAATA	TGCTTATGGT	ATTTAGCTAA	AAGCGGATCA	CATAATTTTT	GAGGGGTGAA	360
TCTGTTTGGC	AGGTCAAGTT	GTCCAATATG	GAAGACATCG	TAAACGTAGA	AACTACGCGA	420
GAATTTTACA	AGTATTAGAA	TTACCAAACCT	TAATAGAAAT	TCAAATAAA	TCTTACGAGT	480

	CTGGTAATTT GTCATTAGAG TTTGTGGATT ACCGTTTAGG AGAACCAAAA TATGATTTAG	600
	AAGAATCTAA AAACCGTGAC GCTACTTATG CTGCACCTCT TCGTGTA AAA GTGCGTCTAA	660
5	TCATTAAAGA AACAGGAGAA GTTAAAGAAC AAGAAGTCTT TATGGGTGAT TTCCCATTA	720
	TGACTGATAC AGGTACGTTT GTTATCAATG GTGCAGAACG TGTAAATCGTA TCTCAATTAG	780
	TTGTTTCACC ATCCGTTTAT TTCAATGAAA AAATCGACAA AAATGGTCGT GAAAACTATG	840
10	ATGCAACAAT TATTCCAAAC CGTGGTGCAT GGTTAGAATA TGAAACAGAT GCTAAAGATG	900
	TTGTATACGT ACGTATTGAT AGAACACGTA AACTACCATT AACAGTATTG TTACGTGCAT	960
	TAGGTTTCTC AAGCGACCAA GAAATGTGTG ACCTTTTAGG TGACAATGAA TATTTACGTA	1020
15	ATACTTTAGA GAAAGACGGC ACTGAAAACA CTGAACAAGC GTTATTAGAA ATCTATGAAC	1080
	GTTTACGTCC AGGTGAACCA CCAACTGTTG AAAATGCTAA AAGTCTATTG TATTCACGTT	1140
20	TCTTTGATCC AAAACGCTAT GACTTAGCAA GCGTGGGTCG TTATAAAACA AACAAAAAT	1200
	TACATTIAAA ACATCGTTTA TTTAATCAAA AATTAGCTGA GCCAATTGTA AATACTGAAA	1260
	CTGGTGAAAT TGTAAGTTGAA GAAGGTACAG TGCTTGATCG TCGTAAATC GACGAAATCA	1320
25	TGGATGTACT TGAATCAAAT GCAAACAGCG AAGTGTGTTGA ATTGCATGGT AGCGTTATAG	1380
	ACGAGCCAGT AGAAATTCAA TCAATTAAAG TATATGTTCC TAACGATGAT GAAGGTCGTA	1440
	CGACAACTGT AATTGGTAAT GCTTCCCTG ACTCAGAAGT TAAATGCATT ACACCAGCAG	1500
30	ATATCATTGC TTCAATGAGT TACTTCTTTA ACTTATTAAG CGGTATTGGA TATACAGATG	1560
	ATATTGACCA TTTAGGTAAC CGTCGTTTAC GTTCTGTAGG TGAATTACTA CAAAACCAAT	1620
	TCCGTATCGG TTTATCAAGA ATGGAAAGAG TTGTACGTGA AAGAATGTCA ATTCAAGATA	1680
35	CTGAGTCTAT CACACCTCAA CAATTAATTA ATATTCGACC TGTTATTGCA TCTATTAAAG	1740
	AATCTTTGG TAGCTCTCAA TTATCACAAT TCATGGACCA AGCAAACCCA TTAGCTGAGT	1800
40	TAACGCATAA ACGTCGTCTA TCAGCATTAG GACCTGGTGG TTTAACACGT GAACGTGCTC	1860
	AAATGGAAGT ACGTGACGTT CACTACTCTC ACTATGGCCG TATGTGTCCA ATTGAAACAC	1920
	CTGAGGGACC AAACATTGGA TTGATTAACT CATTATCAAG TTATGCACGT GTAAATGAAT	1980
45	TCGGCTTTAT TGAAACACCA TATCGTAAAG TTGATTTAGA TACACATGCT ATCACTGATC	2040
	AAATTGACTA TTTAACAGCT GACGAAGAAG ATAGCTATGT TGTAGCACAA GCAAACCTA	2100
	AATTAGATGA AAATGGTCGT TTCATGGATG ATGAAGTTGT ATGTCGTTTC CGTGGTAACA	2160
50	ATACAGTTAT GGCTAAAGAA AAAATGGATT ATATGGATGT ATCGCCGAAG CAAGTTGTTT	2220
	CAGCAGCGAC AgcATGTATT CCATTCTTAG AAAATGATGA CTCAAACCGT GCATTGATGG	2280

55

	CAGGTATGGA ACACGTTGCA GCACGTGATT CTGGTGCGGC TATTACAGCT AAGCACAGAG	2400
	GTCGTGTTGA ACATGTTGAA TCTAATGAAA TTCTTGTTTCG TCGTCTAGTT GAAGAGAACG	2460
5	GCGTTGAGCA TGAAGGTGAA TTAGATCGCT ATCCATTAGC TAAATTTAAA CGTTCAAACT	2520
	CAGGTACATG TTACAACCAA CGTCCAATCG TTGCAGTTGG AGATGTTGTT GAGTATAACG	2580
10	AGATTTTAGC AGATGGACCA TCTATGGAAT TAGGAGAAAT GGCATTAGGT AGAAACGTAG	2640
	TAGTTGGTTT CATGACTTGG GACGGTTACA ACTATGAGGA TGCCGTTATC ATGAGTGAAA	2700
	GACTTGTGAA AGATGACGTG TATACTTCTA TTCATATTGA AGAGTATGAA TCAGAAGCAC	2760
15	GTGATACTAA GTTAGGACCT GAAGAAATCA CAAGAGATAT TCCTAATGTT TCTGAAAGTG	2820
	CACTTAAGAA CTTAGACGAT CGTGGTATCG TTTATATTGG TGCAGAAGTA AAAGATGGAG	2880
	ATATTTTAGT TGGTAAAGTA ACGCCTAAAG GTGTAAGTGA GTTAACTGCC GAAGAAAGAT	2940
20	TGTTACATGC AATCTTTGGT GAAAAAGCAC GTGAAGTTAG AGATACTTCA TTACGTGTAC	3000
	CTCACGGCGC TGGCGGTATC GTTCTTGATG TAAAAGTATT CAATCGTGAA GAAGGCGACG	3060
	ATACATTATC ACCTGGTGTG AACCAATTAG TACGTGTATA TATCGTTCAA AAACGTAAAA	3120
25	TTCATGTTGG TGATAAGATG TGTGGTCGAC ATGGTAACAA AGGTGTCATT TCTAAGATTG	3180
	TTCCTGAAGA AGATATGCCT TACTTACCAG ATGGACGTCC GATCGATATC ATGTTAAATC	3240
	CTCTTGGTGT ACCATCTCGT ATGAACATCG GACAAGTATT AGAGCTACAC TTAGGTATGG	3300
30	CTGCTAAAAA TCTTGGTATT CACGTTGCAT CACCAGTATT TGACGGTGCA AACGATGACG	3360
	ATGTATGGTC AACAAATGAA GAAGCTGGTA TGGCTCGTGA TGGTAAAACT GTACTTTATG	3420
35	ATGGACGTAC AGGTGAACCA TTCGATAACC GTATTTCAGT AGGTGTAATG TACATGTTGA	3480
	AACTTGCGCA CATGGTTGAT GATAAATTAC ATGCGCGTTC AACAGGACCA TATTCAC TTG	3540
	TTACACAACA ACCACTTGGC GGTAAAGCGC AATTCGGTGG ACAACGTTTTT GGTGAGATGG	3600
40	AGGTATGGGC ACTTGAAGCA TATGGTGCTG CATAACATT ACAAGAAATC TTAAC TTACA	3660
	AATCCGATGA TACAGTAGGA CGTGTGAAAA CATACGAGGC TATTGTTAAA GGTGAAAACA	3720
	TCTCTAGACC AAGTGTTCGA GAATCATTCC GAGTATTGAT GAAAGAATTA CAAAGTTTAG	3780
45	GTTTAGATGT AAAAGTTATG GATGAGCAAG ATAATGAAAT CGAAATGACA GACGTTGATG	3840
	ACGATGATGT TGTAGAACGC AAAGTAGATT TACAACAAAA TGATGCTCCT GAAACACAAA	3900
	AAGAAGTTAC TGATTAATAC GCAATTTACA AAACAGGCAA AAAGATACTA AGCTGAATTT	3960
50	TATTGATGAT TCAGTTTAGT ACTTTAAGCC ATTTTAAATA AATGCAAATC AATCAAATAG	4020
	CACAGCTAAT CTAAATTGAA GGAGGTAGGC TCCTTGATTG ATGTAAATAA TTTCCATTAT	4080

55

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55

AAACCTGAAA CAATCAACTA CCGTACATTA AAACCTGAAA AAGATGGTCT ATTCTGTGAA 4200
 AGAATTTTCG GACCTACAAA AGACTGGGAA TG TAGTTGTG GTAAATACAA ACGTGTTCGC 4260
 TACAAAGGCA TGGTCTGTGA CAGATGTGGA GTTGAAGTAA CTAAATCTAA AGTACGTCGT 4320
 GAAAGAATGG GTCACATTGA ACTTGCTGCT CCAGTTTCTC ACATTGGTA TTTCAAAGGT 4380
 ATACCAAGTC GTATGGGATT ATTACTTGAC ATGTCACCAA GAGCATTAGA AGAAGTTATT 4440
 TACTTTGCTT CTTATGTTGT TG TAGATCCA GGTCCAAC TG GTTTAGAAAA GAAACTTTTA 4500
 TTATCTGAAG CTGAATTCAG AGATTATTAT GATAAATACC CAGGTCAATT CGTTGCAAAA 4560
 ATGGGTGCAG AAGGTATTAA AGATTTACTT GAAGAGATTG ATCTTGACGA AGAACTTAAA 4620
 TTGTTACGCG ATGAGTTGGA ATCAGCTACT GGTCAAAGAC TTACTCGTGC AATTAAACGT 4680
 TTAGAAGTTG TTGAATCAAT CCGTAATTCA GGTAACAAAC CTTTCATGGAT GATTTTAGAT 4740
 GTACTTCCAA TCATCCCACC AGAAATTCGT CCAATGGTTC AATTAGATGG TGGACGATTT 4800
 GCAACAAGTG ACTTAAACGA CTTATACCGT CGTGTAATTA ATCGAAATAA TCGTTTGAAA 4860
 CGTTTATTAG ATTTAGGTGC ACCTGGTATC ATCGTTCAAA ACGAAAAACG TATGTTACAA 4920
 GAAGCCGTTG ACGCTTTAAT TGATAATGGT CGTCGTGGTC GTCCAGTTAC TGGCCCAGGT 4980
 AACCGTCCAT TAAAATCTTT ATCTCATATG TTAAAAGGTA AACAAAGGTCG TTTCCGTCAA 5040
 AACTTACTTG GTAAACGTGT TGACTATTCA GGACGTTGAG TTATTGCAGT AGGTCCAAGC 5100
 TTGAAAATGT ACCAATGTGG TTTACCAAAA GAAATGGCAC TTGAACTATT TAAACCATTC 5160
 GTAATGAAAG AATTAGTTCA ACGTGAAATT GCAACTAACA TTAAAAATGC GAAGAGTAAA 5220
 ATCGAACGTA TGGATGATGA AGTTTGGGAC GTATTGGAAG AAGTAATTAG AGAACATCCT 5280
 GTATTACTTA ACCGTGCACC AACACTTCAT AGACTTGGTA TTCAAGCATT TGAACCAACT 5340
 TTAGTTGAAG GTCGTGCGAT TCGTCTACAT CCACTTGTA CAACAGCTTA TAACGCTGAC 5400
 TTTGACGGTG ACCAAATGGC GGTTACGTT CCTTTATCAA AAGAGGCACA AGCTGAAGCA 5460
 AGAATGTTGA TGTTAGCAGC ACAAACATC TTGAACCCTA AAGATGGTAA ACCTGTAGTT 5520
 ACACCATCAC AAGATATGGT ACTTGTAAC TATTACCTTA CTTTAGAAAAG AAAAGATGCA 5580
 GTAAATACAG GCGCAATCTT TAATAATACA AATGAAGTAT TAAAAGCATA TGCAAATGGC 5640
 TTTGTACATT TACACACTAG AATTGGTGTA CATGCAAGTT CGTTCAATAA TCCAACATTT 5700
 ACTGAAGAAC AAAACAAAAA GATTCTTGCT ACGTCAGTAG GTAAAATTAT ATTCAATGAA 5760
 ATCATTCCAG ATTCATTTGC TTATATTAAT GAACCTACGC AAGAAAACTT AGAAAGAAAG 5820
 ACACCAAACA GATATTTTAT CGATCCTACA ACTTTAGGTG AAGGTGGATT AAAAGAATAC 5880

	GAAGTATTCA ACAGATTTAG CATCACTGAT ACATCAATGA TGTTAGACCG TATGAAAGAC	6000
	TTAGGATTCA AATTCTCATC TAAAGCTGGT ATTACAGTAG GTGTTGCTGA TATCGTAGTA	6060
5	TTACCTGATA AGCAACAAAT ACTTGATGAG CATGAAAAAT TAGTCGACAG AATTACAAAA	6120
	CAATTCAACC GTGGTTTAAAT CACTGAAGAA GAAAGATATA ATGCAGTTGT TGAAATTTGG	6180
	ACAGATGCAA AAGATCAAAT TCAAGGTGAA TTGATGCAAT CACTTGATAA AACTAACCCA	6240
10	ATCTTCATGA TGAGTGATTC AGGTGCCCCGT GGTAACGCAT CTAACCTTAC ACAGTTAGCA	6300
	GGTATGCGTG GATTGATGGC CGCACCATCT GGTAAGATTA TCGAATTACC AATCACATCT	6360
15	TCATTCCGTG AAGGTTTAACT AGTACTTGAA TACTTCATCT CAACTCACGG TGCACGTAAA	6420
	GGTCTTGCCG ATACAGCACT TAAAACAGCT GACTCAGGAT ATCTTACTCG TCGTCTTGTT	6480
	GACGTGGCAC AAGATGTTAT TGTTTCGTGAA GAAGACTGTG GTACTGATAG AGGTTTATTA	6540
20	GTTTCTGATA TTAAAGAAGG TACAGAAATG ATTGAACCAT TTATCGAACG TATTGAAGGT	6600
	CGTTATTCTA AAGAAACAAT TCGTCATCCT GAAACTGATG AAATAATCAT TCGTCCTGAT	6660
	GAATTAATTA CACCTGAAAT TGCTAAGAAA ATTACAGATG CTGGTATTGA ACAAATGTAT	6720
25	ATTCGCTCAG CATTTACTTG TAACGCACGA CATGGTGTTT GTGAAAAATG TTACGGTAAA	6780
	AACCTTGCTA CTGGTGAAAA AGTTGAAGTT GGTGAAGCAG TTGGTACAAT TGCAGCCCCA	6840
	TCTATCGGTG AACCAGGTAC ACAGCTTACA ATGCGTACAT TCCATACAGG TGGGGTAGCA	6900
30	GGTAGCGATA TCACACAAGG TCTTCCTCGT ATTCAAGAGA TTTTCGAAGC ACGTAACCCt	6960
	AAAGGTCAAG CGGTAATTAC GGAAATCGAA GGTGTCGTAG AAGATATTAA ATTAGCAAAA	7020
	GATAGACAAC AAGAAATTGT TGTTAAAGGT GCTAATGAAA CAAGATCATA CCTTGCTTCA	7080
35	GGTACTTCAA GAATTATTGT AGAAATCGGT CAACCAGTTC AACGTGGTGA AGTATTAAct	7140
	GAAGGTTCTA TTGAACCTAA GAATTACTTA TCTGTTGCTG GATTAAACGC GACTGAAAGC	7200
40	TACTTATTAA AAGAAGTACA AAAAGTTTAC CGTATGCAAG GTGTAGAAAT CGACGATAAA	7260
	CACGTTGAGG TTATGGTTTCG ACAAATGTTA CGTAAAGTTA GAATTATCGA AGCAGGTGAT	7320
	ACGAAGTTAT TACCAGGTTT ATTAGTTGAT ATTCATAACT TTACAGATGC AAATAGAGAA	7380
45	GCATTTAAAC ACCGTAAGCG TCCTGCAACA GCTAAACCAG TATTACTTGG TATTACTAAA	7440
	GCATCACTTG AAACAGAAAG TTTCTTATCT GCAGCATCAT TCCAAGAAAC AACAAGAGTT	7500
	CTTACAGATG CAGCAATTAA AGGTAAGCGT GATGACTTAT TAGGTCTTAA AGAAAACGTA	7560
50	ATTATTGGTA AGTTAATTCC AGCTGGTACT GGTATGAGAC GTTATAGCGA CGTAAAATAC	7620
	GAAAAAACAG CTAAACCAGT TGCAGAAGTT GAATCTCAAA CTGAAGTAAC GGAATAACAA	7680

55

	ATGTTGACGA ATTCTCTTGT TCAATGTTAA TATATTAAAG GTTGATGCAA GCAGAACTTT	7800
	GGAGGATAAA TTATTGTCTA AGGAAAAAGT tGCACGCTTT AACAAACAAC ATTTTGTAGT	7860
5	TGGTCTTAAA GAAACGCTTA AAGCGTTAAA GAAAGATCAA GTTACATCTT TGATTATTGC	7920
	TGAAGACGTT GAAGTATAIT TAATGACTCG CGTGTTAAGC CAAATCAATC AGAAAAATAT	7980
	ACCTGTATCT TTTTTCAAAA GCAAACATGC TTTGGGTAAA CATGTAGGTA TTAACGTCAA	8040
10	TGCGACAATA GTAGCATTGA TTAAATGAGA ATTAGTAAGT GTTTACTTA CTAAATTTTA	8100
	TTTAACCTAA AAATGAACCA CCTGGATGTG TGGGATTAAA AAGTGAAGAG AGGAGGACAT	8160
	ATCACATGCC AACTATTAAAC CAATTAGTAC GTAAACCAAG ACAAAGCAAA ATCAAAAAAT	8220
15	CAGATTCTCC AGCTTTAAAT AAAGGTTTCA ACAGTAAAAA GAAAAAATTT ACTGACTTAA	8280
	ACTCACCACA AAAACGTGGT GTATGTACTC GTGTAGGTAC AATGACACCT AAAAAACCTA	8340
20	ACTCAGCGTT ACGTAAATAT GCACGTGTGc gTtTATCAAA CAACATCGAA ATTAACGCAT	8400
	ACATCCCTGG TATCGGACAT AACTTACAAG AACACAGTGT TGTACTTGTA CGTGGTGGAC	8460
	GTGTAAAAGA CTTACCAGGT GTGCGTTACC ATATTGTACG TGGAGCACTT GATACTTCAG	8520
25	GTGTTGACGG ACGTAGACAA GGTCGTTTCA TATACGGAAC TAAGAAACCT AAAAACTAAG	8580
	AATTTAGTTT TTAATTAAAT CTTAACTTA AAATATTTAA TATAAGGAAG GGAGGATTTA	8640
	CATTATGCCT CGTAAAGGAT CAGTACCTAA AAGAGACGTA TTACCAGATC CAATTCATAA	8700
30	CTCTAAGTTA GTAACATAAT TAATTAACAA AATTATGTTA GATGGTAAAC GTGGAACAGC	8760
	ACAAAGAATT CTTTATTCAG CATTTCGACCT AGTTGAACAA CGCAGgtTCG TGATGCATTA	8820
	GAAGTATTCG AAGAAGCAAT CAACAACATT ATGCCAGTAT TAGAAGTTAA AGCTCGTCGC	8880
35	GTAGGTGGTT CTAACATCA AGTACCAGTA GAAGTTCGTC CAGAGCGTCG TACTACTTTA	8940
	GGTTTACGTT GGTTAGTTAA CTATGCACGT CTTCTGTGGT AAAAAACGAT GGAAGATCGT	9000
	TTAGCTAACG AAATTTTAGA TGCAGCAAAT AATACAGGTG GTGCCGTTAA GAAACGTGAG	9060
40	GACACTCACA AAATGGCTGA AGCAAACAAA GCATTTGCTC ACTACCGTTG GTAAGATAAA	9120
	AGCTTTTACC CTGAGTGTGT TCTATATTAA TGAATTTTCA TTAAGCGTTC ATGCTTAGGG	9180
45	CATCGCCATA TCTATCGTAT TTATTCAGTA ATATAAAGT GAAGGAGAAA AAATACATGG	9240
	CTAGAGAATT TTCATTAGAA AAAACTCGTA ATATCGGTAT CATGGCTCAC ATTGATGCTG	9300
	GTAAAACGAC TACGACTGAA CGTATTCTTT ATTACACTGG CCGTATCCAC AArGknGGTG	9360
50	AAaCACACGA AGGTGCTTCA CAAATGGACT GGATGGAGCA AGAACAAGAC CGTGGTATTA	9420
	CTATCACATC TGCTGCAACA ACAGCAGCTT GGGAAGGTCA CCGTGTAAC ATTATCGATA	9480

55

	CAGTTACAGT	ACTTGATGCA	CAATCAGGTG	TTGAACCTCA	AACTGAAACA	GTTTGGCGTC	9600
	AGGCTACAAC	TTATGGTGTT	CCACGTATCG	TATTTGTAAA	CAAAATGGAC	AAATTAGGTG	9660
5	CTAACTTCGA	ATACTCTGTA	AGTACATTAC	ATGATCGTTT	ACAAGCTAAC	GCTGCTCCAA	9720
	TCCAATTACC	AATTGGTGCG	GAAGACGAAT	TCGAAGCAAT	CATTGACTTA	GTTGAAATGA	9780
	AATGTTTCAA	ATATACAAAT	GATTTAGGTA	CTGAAATTGA	AGAAATTGAA	ATTCCTGAAG	9840
10	ACCACTTAGA	TAGAGCTGAA	GAAGCTCGTG	CTAGCTTAAT	CGAAGCAGTT	GCAGAACTA	9900
	GCGACGAATT	AATGGAAAAA	TATCTTGGTG	ACGAAGAAAT	TTCAGTTTCT	GAATTAAAAG	9960
	AAGCTATCCG	CCAAGCTaCt	AcTAACGTAG	AAATTCTACCC	AGTACTTTGT	GGTACAGCTT	10020
15	TCAAAAACAA	AGGTGTTCAA	TTAATGCTTG	ACGCTGTAAT	TGATTACTTA	CCTTCACCAC	10080
	TAGACGTTAA	ACCAATTATT	GGTCACCGTG	CTAGCAACCC	TGAAGAAGAA	GTAATCGCGA	10140
20	AAGCAGACGA	TTCAGCTGAA	TTGCTGTCAT	TAGCGTTCAA	AGTTATGACT	GACCCTTATG	10200
	TTGGTAAATT	AACATTCTTC	CGTGTGTATT	CAGGTACAAT	GACATCTGGT	TCATACGTTA	10260
	AGAACTCTAC	TAAAGGTAAA	CGTGAACGTG	TAGGTCGTTT	ATTACAAATG	CACGCTAACT	10320
25	CACGTCAAGA	AATCGATACT	GTATACTCTG	GAGATATCGC	TGCTGCGGTA	GGTCTTAAAG	10380
	ATACAGGTAC	TGGTGATACT	TTATGTGGTG	AGAAAAATGA	CATTATCTTG	GAATCAATGG	10440
	AATTCCCAGA	GCCAGTTATT	CACTTATCAG	TAGAGCCAAA	ATCTAAAGCT	GACCAAGATA	10500
30	AAATGACTCA	AGCTTTAGTT	AAATTACAAG	AAGAAGACCC	AACATTCCAT	GCACACACTG	10560
	ACGAAGAAAC	TGGACAAGTT	ATCATCGGTG	GTATGGGTGA	GCTTCACTTA	GACATCTTAG	10620
	TAGACCGTAT	GAAGAAAGAA	TTCAACGTTG	AATGTAACGT	AGGTGCTCCA	ATGGTTTCAT	10680
35	ATCGTGAAAC	ATTCAAATCA	TCTGCACAAG	TTCAAGGTAA	ATTCTCTCGT	CAATCTGGTG	10740
	GTCGTGGTCA	ATACGGTGAT	GTTACATTG	AATTCACACC	AAACGAAACA	GGCGCAGGTT	10800
40	TCGAATTCTA	AAACGCTATC	GTTGGTGGTG	TAGTTCCTCG	TGAATACATT	CCATCAGTAG	10860
	AAGCTGGTCT	TAAAGATGCT	ATGGAAAATG	GTGTTTATGC	AGGTTATCCT	TTAATTGATG	10920
	TTAAAGCTAA	ATTATATGAT	GGTTCATACC	ATGATGTCGA	TTCATCTGAA	ATGGCCTTCA	10980
45	AAATTGCTGC	ATCATTAGCA	CTTAAAGAAG	CTGCTAAAAA	ATGTGATCCT	GTAATCTTAG	11040
	AACCAATGAT	GAAAGTAACT	ATTGAAATGC	CTGAAGAGTA	CATGGGTGAT	ATCATGGGTG	11100
	ACGTAACATC	TCGTCGTGGA	CGTGTGTGATG	GTATGGAACC	TCGTGGTAAT	GCACAAGTTG	11160
50	TTAATGCTTA	TGTACCACTT	TCAGAAATGT	TCGGTTATGC	AACATCATTA	CGTTCAAACA	11220
	CTCAAGGTCG	CGGTACTTAC	ACTATGTACT	TCGATCACTa	TGCTGAAGTT	CCaAAATCaA	11280

55

	GCCTAGGTTA AAATACAAGG TGAGCTTAAA TGTAAGCTAT CATCTTTATA GTTTGATTTT	11400
	TTGGGGTGAA TGCATTATAA AAGAATTGTA AAATTCTTTT TGCATCGCTA TAAATAATTT	11460
5	CTCATGATGG TGAGAAACTA TCATGAGAGA TAAATTTAAA TATTATTTTT AATTAGAATA	11520
	GGAGAGATTT TATAATGGCA AAAGAAAAAT TCGATCGTTC TAAAGAACAT GCCAATATCG	11580
	GTAATATCGG TCACGTTGAC CATGGTAAAA CAACATTAAC AGCAGCAATC GCTACTGTAT	11640
10	TAGCAAAAAA TGGTGACTCA GTTGACACAAT CATATGACAT GATTGACAAC GCTCCAGAAG	11700
	AAAAAGAACG TGGTATCACA ATCAATACCT CTCACATTGA GTACCAAACG GACAAACGTC	11760
	ACTACGCTCA CGTTGACTGC CCAGGACACG CTGACTACGT TAAAAACATG ATCACTGGTG	11820
15	CTGCTCAAAT GGACGGCGGT ATCTTAGTAG TATCTGCTGC TGACGGTCCA ATGCCACAAA	11880
	CTCGTGAACA CATTCTTTTA TCACGTAACG TTGGTGTAAC AGCATTAGTA GTATTCTTAA	11940
20	ACAAAGTTGA CATGGTTGAC GATGAAGAAT TATTAGAATT AGTAGAAATG GAAGTTCGTG	12000
	ACTTATTAAG CGAATATGAC TTCCCAGGTG ACGATGTACC TGTAATCGCT GGTTGAGCAT	12060
	TAAAAGCTTT AGAAGGCGAT GCTCAATACG AAGAAAAAAT CTTAGAATTA ATGGAAGCTG	12120
25	TAGATACTTA CATTCCAACG CCAGAACGTG ATTCTGACAA ACCATTCTATG ATGCCAGTTG	12180
	AGGACGTATT CTCAATCACT GGTCTGGTGA CTGTTGCTAC AGGCCGTGTT GAACGTGGTC	12240
	AAATCAAAGT TGGTGAAGAA GTTGAAATCA TCGGTTTACA TGACACATCT AAAACAACG	12300
30	TTACAGGTGT TGAAATGTTC CGTAAATTAT TAGACTACGC TGAAGCTGGT GACAACATTG	12360
	GTGCATTATT ACGTGGTGTT GCTCGTGAAG ACGTACAACG TGGTCAAGTA TTAGCTGCTC	12420
	CTGGTTCAAT TACACCACAT ACTGAATTCA AAGCAGAAGT ATACGTATTA TCAAAGACG	12480
35	AAGGTGGACG TCACACTCCA TTCTTCTCAA ACTATCGTCC ACAATTCTAT TTCCGTACTA	12540
	CTGAAGTAAC TGGTGTGTT CACTTACCAG AAGGTACTGA AATGGTAATG CCTGGTGATA	12600
	ACGTTGAAAT GACAGTAGAA TTAATCGCTC CAATCGCGAT TGAAGACGGT ACTCGTTTCT	12660
40	CAATCCGTGA AGGTGGACGT ACTGTAGGAT CAGGCGTTGT TACTGAAATC ATTAAATAAT	12720
	TTCTAATTTT TTAGATTTTA TATAAAAAGA AGATCCCTCA ATCGAGGGG CTTTTTTTAA	12780
45	TGTGTAAATT TTGTAATGGC TATTCGATTT AGAAGAACAA TAATTGATGA AAGACTGACT	12840
	AATAAACTT ATAAGTGATA ATACTGTTTA AATAAAATTG TTGAGTCTTG GACATTGTAA	12900
	AATGCTCCCT TCAAAGTTTT CATTTTTTCa ATGTCTACTT TGAAGGGAGC ATTTTCATTAG	12960
50	TTTATGTCTC AGATTCATAT CTTTCAATTA ATTTAAATGC TTAATTTGTT TTAAATACTT	13020
	GCTCTAATTC TATGATTTTT AAAAATACAG CTACAGCGTA TTTTAATGAT TTTTCATCAA	13080

55

	TCAGAAAGAA TGCACCTGGT CGTACTTTCA AATAATGTGA AAAATCTTCT CCAATCATCA	13200
	TTAAATCTGA TTCATTAAAG CGTACATGTA AGTCATTGTG TGCTTCTTTA ATAACCTTGAT	13260
5	ATGCTTTCTC GTTATTATGG ACAGGCAAAT ACCCTTTAAT ATAATTCAAA TCATAGTTAA	13320
	TATCATTTGC TATTGCTAAA CCTGTAGAA GCTTATCCAT TTTGTCCATT ACATGATTCT	13380
	GTATATCTGA ATCGAAAGTT CTAACGTAC CTTTACAAAA TGCTTGATCA GGAATAACGC	13440
10	TATCTGTGGT GCCTGCTTGA ATCATTCCAA ATGAAAGTAC AGCTTGTTTA ACTGGATCGA	13500
	TCGTACGTGA AATTATTTTT TGTGCACTTA AAATGAACTC TGCCATGATT ACTATTGGGT	13560
	CAATGGTTTC ATGAGGTTTG GCACCATGAC CACCACGACC TTAAATGTG ACGCTAAATT	13620
15	CATCTGGAGA GGCCATGATT GCGCCGCAC GTGAATGAAT AGTTCCAGTA GGATAACCAC	13680
	TCCATAAATG TGTACCGTAA ATTCTATCTA CATTTTCCAG ACATCCAGCA TCTATCATTT	13740
20	CTTGAGAACC ACCTGGCATG ATTTCTTCAC CGTACTGGAA TATTAATACA ACATTACCTT	13800
	CTAATAAATG TTTATGTTCA TCTAAAATCT CTGCTACAGT AAGTAAATT GCTGTATGAC	13860
	CATCATGCCC ACACGCATGC ATACATCCTG GATTTTTAGA CTTATAAGGC ACATCGTTTA	13920
25	ATTCTCGAC AGGTAACGCA TCAAAGTCAG CTCTTAATGC AATGGTAGGT CCTGTGCCCA	13980
	AGCCTTTAAA TGTGGCTTTG ATACCATTGC GGCCGATAGG AGTTTCAATA TCACAAGATA	14040
	ACTGGCTTAA TTGGTTAACA ATATAATCAT GTGTTTGAAA TTCTTCAAAA GATAACTCAG	14100
30	GATATTGGTG TAAATAACGT CTGAGTTGAA TTGTTTTATT TTCTTTATTA TTTGCTAGTT	14160
	GGAACCAATC TAACACCCTT ATCACTACTT TCTAAAATAA TGTTTATAGT ATAACATTTT	14220
	ATGAAATTAT CGTACTAAAT GATTGCTTTG AGATATTTTA TCTATGAATG ATAAGGCTTT	14280
35	CAAGTTATGT AGAATTACTG TATGATAAAG GTATTACCAA ACAATACTTA AGGGGGATTA	14340
	TATACTGTGG TTCAATCATT ACATGAGTTT TTAGAGGAAA ATATAAATTA TCTAAAAGAA	14400
40	AATGTTTGT ATAATGAAAT AGATACAATT GAAGGTGCAA ACGGACCAGA AATCAAAATC	14460
	AATGGGAAAT CATACATTAA CTTATCTTCA AATAATTATT TAGGACTAGC AACAAATGAA	14520
	GATTTGAAAT CaGctGCAAA AGCAGCTATT GATACACATG GTGTAGGTGC AGGCGCTGTT	14580
45	CGTACAATCA ATGGTACATT AGATTTACAC GACGAATTAG AAGAAACACT AGCAAAATTT	14640
	AAAGGAACAG AAGCTGCAAT AGCTTATCAA TCAGGATTTA ATTGTAATAT GGCTGCTATT	14700
	TCAGCTGTCA TGAATAAAAA TGATGCTATT TTATCAGATG AGCTTAATCA TGCATCAATT	14760
50	ATTGATGGAT GTCGCTTATC TAAAGCTAAA ATTATTCGAG TTAACCATTC AGACATGGAT	14820
	GATTTACGTG CGAAAGCAAA AGAAGCAGTT GAATCAGGTC AATACAATAA AGTGATGTAT	14880

55

ATTGCAGAAG AATTTGGTTT ATTAACCTAT GTTGACGACG CTCATGGTTC AGGTGTTATG 15000
 GGTAAGGCG CTGGTACGGT TAAACATTTT GGTTTACAAG ATAAAATCGA TTTCCAAATA 15060
 5 GGTACGCTTT CTAAAGCAAT TGGTGTGCGT GCGGGTTATG TAGCAGGTAC AAAAGAGTTA 15120
 ATAGATTGGT TAAAAGCACA ATCACGACCA TTCTTATTCT CTACATCATT AGCACCTGGG 15180
 GATACCAAAG CAATAACTGA AGCAGTTAAA AAGTTAATGG ATTCAACTGA ATTACATGAT 15240
 10 AAATTATGGA ACAATGCACA ATATTTAAAA AATGGATTGT CAAAATTAGG ATATGATACA 15300
 GGTGAGTCAG AAACCTCAAT TACACCAGTA ATTATTGGTG ATGAAAAAAC AACTCAAGAA 15360
 TTTAGTAAGC GTTTAAAAGA CGAAGGTGTC TATGTGAAAT CTATCGTTTT CCCAACAGTA 15420
 15 CCAAGAGGTA CAGGACGTGT AAGAAATATG CCTACAGCTG CACATACAAA AGACATGTTA 15480
 GATGAAGCAA TTGCGGCTTA TGAAAAAGTA GGAAAAAGAA TGAAGTTGAT TTAATATTTA 15540
 20 TTTATTCCCA CGGCAAATAT TGTGCTGGGC TTTTTTAAAT GTTTAGTTTA TTAACAGT 15598

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 661 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

AAGTAAATCA ACTTACTGGG ATAAGAATAA AGGCGATTAT AGTAACAAGT TGATTTTATT 60
 CGAAAAACAT TTTGAACCGG TTCTGGGTAT CAAGATGCAA CATAGTGGAG GTCATAGCTT 120
 35 TGGCCACACG ATTATTACGA TTGAAAGTCA AGGAGATAAA GCAGTTCATA TGGGTGATAT 180
 ATTCCCAACT ACTGCACATA AAAATCCTCT ATGGGTAACG GCATATGATG ATTATCCTAT 240
 GCAATCGATT CGTGAAAAAG AACGCATGAT ACCATATTTT ATTCAGCAAC AATATTGGTT 300
 40 CTTGTTTTAT CATGATGAAA ACTACTTTGC TGTAATAATAC AGCGATAATG GTGAAAACAT 360
 AGATGCATAT ATTTTACGTG AAACATTAGT TGATAATAAC TAAAATAAAG ATGTATTACT 420
 AAACAAATTT TCAAAAATAA AAAATTGAGC CACATCCAAT CTTACTAATT AGGGTGTGGC 480
 TCATTTTTAA GTTTTACgAT CCAATCAAA TATGGaTAAA ATTCgTATTA ACGCTCTACa 540
 ATGtTAATGA CTTCAACAGT ATATGCATCT GCATAAAAAT CATAATGAAT ATTTTGACCA 600
 50 TTTTTAATAG TTGTAATTCC ACCTTGATAA ACTAAACGGT ATTTATCAGT TTCAGGATGA 660
 A 661

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5738 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

10 GCAGACGGTA CAGCAGTTAA AGTCGCACCA AaACTGTagT GAATcTAATC GGTGcATTCT 60
 TTTTAGGATT AGTTGTGCGG CTTATATATA TCTTCTTCAA AGTAATTTTC GATAAGCGAA 120
 15 TTAAAGATGA AGAAGATGTA GAGAAAGAAT TAGGATTGCC TGTATTGGGT TCAATTCAAA 180
 AATTTAATTA AGGATGGTTG CTACTTATGT CAAAAAAGGA AAATACGACA ACAACACTAT 240
 TTGTATATGA AAAACCAAAA TCAACAATTA GTGAAAAGTT TCGAGGTATA CGTTCAAACA 300
 20 TCATGTTTTTC AAAAGCAAAT GGTGAAGTAA AGCGCTTATT GGTTACTTCT GAAAAGCCTG 360
 GTGCAGGTAA AAGTACAGTT GTATCGAATG TAGCGATTAC TTATGCACAA GCAGGCTATA 420
 AGACATTAGT TATTGATGGC GATATGCGTA AgcCAACACA AACTATATT TTTAATGAGC 480
 25 AAAATAATAA TGGACTATCA AGCTTAATCA TTGGTCGAAC GACTATGTCA GAAGCAATTA 540
 CGTCGACAGA AATTGAAAAT TTAGATTGTC TAACAGCTGG CCCTGTACCT CCAAATCCAT 600
 CTGAGTTAAT TGGGTCTGAA AGGTTCAAAG AATTAGTTGA TCTGTTAAT AAACGTTACG 660
 30 ACATTATTAT TGTCGATACA CCGCCAGTTA ATACTGTGAC TGATGCACAA CTATATGCGC 720
 GTGCTATTAA AGATAGTCTG TTAGTAATTG ATAGTGAAAA AAATGATAA AATGAAGTTA 780
 AAAAAGCAAA AGCACTTATG GAAAAAGCAG GCAGTAACAT TCTAGGTGTC ATTTTGAACA 840
 35 AGACAAAGGT CGATAAATCT TCTAGTTATT ATCACTATTA TGGAGATGAA TAAGTATGAT 900
 TGATATTTCAT AACCATATAT TGCCTAATAT CGATGACGGT CCGACAAATG AAACAGAGAT 960
 GATGGATCTT TTAAAACAAG CGACAACACA AGGTGTTACA GAAATCATTG TAACATCACA 1020
 40 TCACTTACAT CCTCGATATA CCACACCTAT AGAAAAAGTG AAATCATGTT TAAACCATAT 1080
 TGAAAGCTTA GAGGAAGTAC AAGCACTAAA TCTAAAGTTT TATTATGGTC AGGAAATAAG 1140
 45 AATTACCGAT CAAATCCTTA ATGATATTGA TCGAAAAGTT ATTAACGGTA TTAATGATTC 1200
 ACGCTATTTA CTAATAGAAT TTCCATCAAA TGAAGTTCCA CACTATACTG ATCAATTATc 1260
 TTTCGAATcA CAGAGTAAAG GCTTTGTACC GATTATTGCA CATCCAGAGC GGAATAAAGC 1320
 50 AATAAGTCAA AACCTTGACA TACTATACGA TTTAATTAAC AAAGGTGCTT TAAGTCAAGT 1380
 GACAACGGcG TCATTAGCGG GTATTTCGG TAAAAAAATT AGAAAATTAG CAATTCAAAT 1440

55

	GTTCTTAATG AAAGACTTAT TTAATGATAA GAAATTACGT GATTATTATG AAGATATGAA	1560
	CGGATTTTATT AGTAATGCGA AGTTAGTTGT TGATGATAAA AAAATTCCTA AACGAATGCC	1620
5	ACAACAAGAT TATAAACAGA AAAGATGGTT TGGGTATATA ACAGCAAATG AGGGGTTTTA	1680
	TGGCACATTT ATCTGTGAAA TTGCGGCTTT TAATACTAGC ATTAATCGAT TCACTGATAG	1740
	TGACATTTTC AGTATTCGTA AGTTATTACA TTTTAGAACC GTATTTCAAA ACATATTCTG	1800
10	TCAAATTATT AATATTGGCA GCTATATCAC TATTCATATC GCATCATATT TCaGCATTTA	1860
	TTTTTAATAT GTATCATCGA GCGTGGGAAT ATGCCAGTGT GAGTGAATTG ATTTTAATTG	1920
	TTAAAGCTGT GACGACATCT ATCGTTATTA CGATGGTGGT CGTGACAATT GTTACAGGCA	1980
15	ATAGACCGTT TTTTAGATTG TATTTAATTA CTTGGATGAT GCACTTGATT TTAATAGGTG	2040
	GCTCAAGGTT ATTTTGGCGT ATTTATCGGA AATACCTTGG AGGTAAGTCA TTTAATAAGA	2100
20	AGCCAACTTT AGTTGTTGGT GCTGGTCAAG CAGGTTCAAT GCTGATTAGA CAAATGTTGA	2160
	AAAGTGACGA AATGAACTT GAACCGGTAT TAGCAGTCGA TGATGACGAA CATAAACGCA	2220
	ATATCACAAT TACTGAGGGT GTAAAAGTCC AAGGTAAAAT TGCGGATATT CCAGAACTAG	2280
25	TGAGGAAATA TAAGATTAAA AAAATCATCA TTGCAATTCC AACTATTGGT CAAGAGCGTT	2340
	TGAAAGAAAT TAATAATATT TGCCATATGG ATGGCGTTGA GTTATTGAAA ATGCCAAATA	2400
	TAGAAGACGT CATGCTGGT GAGTTAGAAG TGAACCAACT TAAAAAGTT GAAGTAGAAG	2460
30	ATTTACTAGG CAGAGATCCT GTTGAATTAG ATATGGATAT GATATCAAAT GAATTGACGA	2520
	ATAAACTAT TTTAGTTACG GGTGCAGGTG GTTCAATAGG ATCAGAAATT TGTAGACAAG	2580
	TTTGTAATTT CTATCCAGAA CGTATTATTC TACTTGGCCA TGGTGAAAAC AGTATTTATT	2640
35	TAATCAATCG TGAATTGCGA AATCGCTTCG GwAAAAATGT TGATATCGTT CCTATTATAG	2700
	CGGATGTGCA AAATAGAGCG CGTATGTTTG AAATTATGGA AACGTATAAA CCATACGCAG	2760
40	TTTATCATGC AGCAGCACAC AAGCACGTGC CGTTAATGGA AGACAACCCT GAAGAAGCAG	2820
	TACGTAATAA TATTTTAGGT ACGAAAAATA CTGCTGAAGC TGCTAAAAAT GCAGAGGTAA	2880
	AGAAATTCGT TATGATTTCT ACGGATAAAG CCGTTAATCC GCCTAATGTC ATGGGAGCTT	2940
45	CAAAGCGAAT TGCAGAAATG ATTATTCAAA GTTTAAATGA TGAAACGCAT CGAACAAAT	3000
	TTGTTGCAGT GAGATTGGT AATGTACTTG GATCGAGAGG ATCTGTGATT CCACTTTTCA	3060
	AAAGTCAAAT TGAAGAAGGT GGGCCAGTTA CTGTGACACA TCCTGAAATG ACACGTTACT	3120
50	TTATGACAAT TCCTGAAGCT TCTAGACTAG TTTTGCAGGC AGGGGCATTA GCAGAAGGTG	3180
	GCGAAGTATT TGTGCTAGAT ATGGGAGAAC CAGTGAAAAT TGTAGATTTG GCACGTAATT	3240

55

	CCGGCGAAAA AATGTTTGAA GAGCTTATGA ATAAAGATGA GGTTCATCCT GAACAAGTAT	3360
	TTGAAAAAAT TTATCGTGGC AAAGTACAAC ATATGAAATG TAATGAAGTT GAAGCGATTA	3420
5	TTCAAGACAT CGTCAATGAC TTTAGTAAAG AAAAAATTAT TAACTATGCC AATGGCAAAA	3480
	AGGGAGATAA TTATGTTTCA TGACAAAATT TTATTAATTA CTGGGGGCAC AGGATCATTC	3540
	GGTAATGCTG TTATGAAACA GTTTTTAGAT TCTAATATTA AAGAAATTCG TATTTTTTCA	3600
10	CGCGATGAGA AAAAAACAAGA TGACATTCGA AAAAAATATA ATAATTCAAA ATTAAAGTTC	3660
	TACATTGGTG ATGTGCGTGA TAGTCAAAGT GTAGAACAG CAATGCGAGA TGTGATTAC	3720
	GTATTCCATG CAGCAGCTTT AAAACAAGTG CCGTCATGTG AATTCCTTCC AGTTGAGGCA	3780
15	GTGAAGACAA ATATTATTGG TACAGAAAAT GTCTTACAAA GTGCTATTCA TCAAAATGTT	3840
	AAAAAAGTCA TATGTTTATC TACAGATAAG GCAGCGTATC CTATTAATGC TAGGGGTATT	3900
20	TCAAAAGCAA TGATGGAAAA AGTATTCGTA GCCAAATCAA GAAATATTCG TAGTGAACAA	3960
	ACGCTTATTT GTGGTACAAG ATACGGTAAT GTGATGGCTT CAAGAGGATC AGTAATACCT	4020
	TTGTTTATCG ACAAATCAA AGCTGGAGAA CCTTTAACGA TTACAGATCC TGATATGACA	4080
25	AGATTTTTAA TGAGCTTAGA AGATGCGGTA GAACTAGTTG TTCATGCATT TAAGCATGCA	4140
	GAGACAGGAG ATATTATGGT TCAAAAAGCA CCAAGCTCAA CGGTAGGGGA TCTTGCGACC	4200
	GCATTATTAG AATTGTTTGA AGCTGATAAT GCAATTGAAA TCATTGGTAC GCGACATGGA	4260
30	GAGAAAAAAG CAGAAACATT GTTGACGAGA GAAGAATACG CACAATGTGA AGATATGGGT	4320
	GATTATTTTA GAGTGCCGGC AGACTCCAGA GATTTAAATT ATAGTAATTA TGTGAAACC	4380
	GGTAACGAAA AGATTACGCA ATCTTATGAA TATAACTCCG ATAATACACA TATTTTAACG	4440
35	GTGGAAGAGA TAAAGAAAA ACTTTTAACA CTAGAATATG TTAGAAACGA ATTGAATGAT	4500
	TATAAAGCTT CAATGAGATA GGAGAGATTG ACGTTGAATA TTGTAATTAC AGGAGCAAAA	4560
40	GGTTTTGTAG GAAAAAAGCTT GAAAGCAGAT TTAAGTTCAA CGACAGATCA TCATATTTTC	4620
	GAAGTACATC GACAAACTAA AGAGGAAGAA TTAGAGTCAG CATTGTTGAA AGCAGACTTT	4680
	GTCGTGCATT TAGCGGGTGT TAATCGACCT GAACATGACA AAGAATTCAG CTTAGGAAAC	4740
45	GTGAGTTATT TAGATCATGT ACTTGATATA TTAAGTAGAA ATACGAAAAA GCCAGCGATA	4800
	TTATTATCGT CTTCAATACA AGCAACACAA GATAATCCTT ATGGTGAGAG TAAGTTGCAA	4860
	GGGGAACAGC TATTAAGAGA GTATGCCGAA GAGTATGGCA ATACGGTTTA TATTTATCGC	4920
50	TGGCCAAATT TATTCGGCAA GTGGTGTAAG CCGAATTATA ACTCAGTGAT AGCAACATTT	4980
	TGTTACAAAA TTGCACGTAA CGAAGAGATT CAAGTTAATG ATCGGAATGT TGAACATAACG	5040

55

ATTGAAAATG GTGTACCTAC AGTACCAAAC GTATTTAAAG TGACATTGGG AGAAATTGTA 5160
 GATTTATTAT ACAAGTTCAA ACAGTCACGT CTCGATCGAA CATTGCCGAA ATTAGATAAC 5220
 5 TTGTTTGAAA AAGATTTGTA TAGTACGTAT TTAAGCTATC TACCTAGTAC aGACTTTAGT 5280
 TAYCCCTTAC TTATGAATGT GGATGATAGG GGTTCCTTTA CAGAATTTAT AAAAACACCG 5340
 GATCGTGGTC AAGTTTCTGT AAATATTTCT AAACCAGGTA TTAATAAGG TAATCACTGG 5400
 10 CATCATACTA AAAACGAAAA ATTTCTAGTC GTATCAGGTA AAGGGGTAAT TCGTTTITAGA 5460
 CATGTTAATG ATGATGAAAT CATTGAATAT TATGTTTCTG GCGACAAAT AGAAGTTGTA 5520
 GACATACCAG TAGGATACAC ACATAATATT GAAAATTTAG GCGACACAGA TATGGTAACT 5580
 15 ATTATGTGGG TGAATGAAAT GTTGATCCA AATCAGCCAG ATACGTATTT CTTGGAGGTA 5640
 TAGCGCATGG aAAAACTGAA rTTAATGACA ATAGTTGGTA CAAGGCCTGA AATCATTCGT 5700
 20 TTATCATCAA CGATTAAAGC ATGTGATCAA TATtTTAA 5738

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9062 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

ATCATCAACA AGAATGATAT TTTTCCCATC TACTATATCT TTTACCGCAG ATAACTTCAC 60
 TCTCACACCT TGCTCACGTA ATTCTTGAGT TGGTTGAATA AATGTTCTTG CAACATATTG 120
 35 ATTTTAACT AGTCCCATTT CATATGGCAA ACCTATTTCT TCAGCATAAC CACTCGCAGC 180
 TGATAGCGAT gAATTGGGTA CACCGATGAC CATATCAGCA TTTACAGGGC TTTCTTGGGC 240
 40 TAATTTTTTA CCAGAAGCTT TACGTACTGC ATGGACATTT TTACCAGCTA TTGTTGAGTC 300
 TGGTCTAGCA AAATAAATAT ATTCCATCGC AGAAATTGCA GTTGTCGTAT GATGTGTATA 360
 AGATTAACT GTAATACCTT TATCGTTAAT CACGACATAT TCACCTGCAT GAATATCTTG 420
 45 AACAAATTCT GCACCTAACA CATCTATTGC ACATGTTTCA CTTGCAAGGA TGTATGTCCC 480
 ATCTTTCATT TTACCTACAA CAAGTGGTCT GATAGCATTT GGATCTACTG CGCCATATAA 540
 CGCATCTTTA GTTAAAATCG CAAATGTAAA ACCGCCTTTA ACTTTTCGCA AACTTTCTTT 600
 50 CAACGCTTCC TCAAAAGTAG GAGCTTTACT TCGACGTATC AAATGCATAA TGACTTCAGT 660
 ATCAGAAGAC GAATGGAAGA TAGCACCTTG TTTTCTTAAA TTCTGACGCA ATGATTAGC 720

55

	CGGTTGAATA TTTTCAATAC CTTTATTACC TGAAGTAGCA TAACGGACGT GACCAATTGC	840
	ATGTTGATAT CCTTTTAATC GTTCCATTG ATCATCTTAA ATCGCTTCAG TTAGTAAGCC	900
5	TAATCCTCGC TCGCCTTTTA ATTCATTTTG ATCAGAAACA ACTATACCTG cACCTTCTTG	960
	ACCACGATGT TGCAAACTAT GAAGTCCCAT ATAtGTTAGT TGCCTGCTt CaGGATGATT	1020
	CCAAATACCA AACACGCCAC ATTCTTCGTT TAATCCTGAG TAGTTAAACA TTGaGCAATT	1080
10	GCCCCtTCCC ATATTTGTTT AATATCTGAA ACATTTTCAC TAATCTCTGT aTATGGTGTT	1140
	GTTACCTTGr aATTATCACT ATCTGTTAAA AGTCCAATTT CTATTGCATT ATCAATATTT	1200
	AAAGTTTAC CTGATTTAAC AGAAACAACA TATCGGCCTT GCGTCTCACT AAACAATTGT	1260
15	GCATTTGTTA TATCTATTGA AGATTTTAAAT CCTAAACCGT AATGCGCACT TAGTTTAGCT	1320
	AAGGTAATCA GTAAGCCACC TTTACCAACT GTTTGAACAT GTGATAATAG TCCTTCACGA	1380
20	ATAGCGGTCT TGATTGATTC ACCTTTTTCA ACTTCTGAAC TCAAATCTAA TGA CTCAAAT	1440
	TCATGATTAA CTTTGCCATA AATTAACTTT TCAAGTTGAC TACCACCAA GTCGTCCTTA	1500
	GTATCACCGA TTAAATATAA TTTATCTCCA ACTTGAGGTT CAAAATCATT TAAATAATTT	1560
25	ACATTTTCAA TCAAACCTAC CATTCCAACA ACTGGTGTG GGAAATAGA AGTACCTTTC	1620
	GTTTCGTTAT ATAAAGATAC ATTACCAGAA ACTACTGGTG TCTTAAGAAT GTCGCATGCT	1680
	TCTGCCATAC CTTTCGTTGA ATCTATCAAC TGTTGATAGA TTTCTTTCTT TTCAGGAGAA	1740
30	CCATAATTTA AACAACTCTGT CATTGCTAAT GGTGTTGCAC CCACGGCAAT TAAATTTCGA	1800
	TAAGCTTCAG CTACTACCAT CTTTCACCT TCATATGGAT TGTATATAC ATAACGCGCT	1860
	TCACCATCAA TTGTTGAAGC AATTGCCTTA TTTGTGCCTT CCACACGTAC TACCGATGCT	1920
35	TGAAGTCCTG GCTTAATTAT CGTATTGGCA CCAACTTGTT GGTCGTATTG ATCATATAAA	1980
	TAGTgTTTTAG ATGCTATAGT CGGATGCTTA AGTAATTTAA AGAAAGTATC TTTAACATCG	2040
40	ATGTGTGTAT AATCATTTTT AGAAGTATTA TAATCTTTTT CTTCTCCTTC TAAAATATAT	2100
	ACAGGTGCTT CATCAGCTAG TGGTTCAACT GGAATGTCAG CATAAACTTC GTCATCATAT	2160
	GTTAAACAA AACGATTTGT ATCTGTAACT TCACCTATAA CAGCACTATC CAATTCGTGC	2220
45	TTATCAAATA AATCTAAGAA TTTTGTTC GTACCTTTTT CAACAACCTAG TAACATACGT	2280
	TCTTGAGTTT CTGAAAGCAT CATTTTATAA GGAGAAATAC CTGGCTCACG TGTGGCACT	2340
	TGTTCTAATC TCAAATGTAA CCCACTACCA CCTTTTGCCG CCATTTCAGA CGATGAAGAT	2400
50	GTTAAACCAG CAGCACCCAT ATCTTGAATA CCAACTAATT CATCAAATGT AATTGCTTCA	2460
	AGTGTTGCTT CCATTAATTT TTTACCTACA AATGGATCAC CGATTTGTAC AGAAGGTCGT	2520

55

	CGACCAGTTT	TCAAACCAAC	ATAAATGACC	GAATTACCTA	CACCTTTTGC	TGTGCCTTTT	2640
	TGAATCATGT	CGTGATTGaT	AACACCAACA	CACATTGCAT	TAACAAGTGG	ATTGCCATCA	2700
5	TAACGTTCAT	CAAATTCGAT	TTCACCAGCA	GTTGTTGGaA	TACCAATGCA	GTTACCATAA	2760
	CCTCCGATAC	CCTTTACAAC	ACCTTTAAGT	AATCTTTGGT	TTTGTTTATT	ATCTAATTCT	2820
	CCAAATCTAA	GACTGTTTAA	CAAATTAATA	GGTCTAGCCC	CAATAGAGAC	AATGTCACGA	2880
10	ATGATTCCAC	CAACGCCTGT	AGCAGCCCCT	TGATATGGTT	CAATTGCTGA	TGGATGATTG	2940
	TGAGACTCTA	CTTTAAATAC	TACGGCTTGA	TTATCACCTA	TATCGACTAC	CCCTGCACCT	3000
	TCACCAGGCC	CCATAAGCAC	ATGGTcACCT	GACGTAGGAA	ATTGCTTTAA	AAACGGTTTA	3060
15	GAATGTTTAT	AAGAGCAATG	TTCACTCCAC	ATAACAGAAA	AGATACCTGT	TTCTGTAAAG	3120
	TTAGGTTGTC	TGCCTAAAAT	ATCGCAAAC	TTTTCATATT	CTTGATCaCT	TAATCCCATA	3180
20	TCTTGATATA	CTTTTTCAAG	TTTAATTTCT	TCAACGCTTG	GTTGATAAAA	TTTAGACATG	3240
	TTGTTCCCTC	CAACTTTTTA	CCATCGCTTC	AAATAATTTT	ACACCACTAT	CAGTACCTAA	3300
	CAACGTTTCT	AAAGCTCTTT	CagGATGtGG	CATCATGCCA	CATACATTGC	CTTTTTTCGTT	3360
25	AACAATTCCT	GCAATATCAT	CATATGAACC	GTTCCGATTA	TTCACATATT	TCAGAATAAT	3420
	TTGATTGTTA	GCTTTTAAAT	GTTGATATAT	TTCATCAGTA	CAATAATAAT	GACCTTCACC	3480
	GTGAGCTACA	GGATATATAA	CTTTTTCACC	TTGTTCATAA	AGATTTGTAA	ATGCCGTTTG	3540
30	ATTATTCACT	ATTTCTAACT	CTTCATTTCT	ACTAATAAAT	AAATGTGAAT	CGTTATGCAA	3600
	TAATGCACCA	GGTAATAAGC	CTATTTCACT	TAAAATTTGA	AACCCATTAC	AAACACCTAA	3660
	TACTGGCTTA	CCTTCAGCTG	CAAGACGTTT	AACTTCCGAA	ATAATCGGsG	CTACACTAGC	3720
35	CATTGCCCCA	GATCTTAAGT	AATCCCCGAA	TGAAAATCCA	CCAGGAATAA	GTACGCCATC	3780
	AAATēCACTT	AGTGATGTTT	CTCTATAATC	TACATATTCC	GCTTCAACAC	CACTTTTAAAT	3840
40	AGCAGCATTA	AACATGTCTC	TATCACAATT	CGAACCTGGA	AAAACAAGAA	CCGCAAATTT	3900
	CATTTTATGC	ATTCTCCTTT	TCATCATCTA	ACACTTTATA	GCTATAITCT	TCAATCACTG	3960
	TATTTGCAAA	CAATTTTTCA	CTTAGAGTTG	TAATAATGTT	GTGTACCTTT	TCATCACTAA	4020
45	CCTCATCCAC	TGTCATATAT	AATACTTTTC	CTACACGAAT	ATCATTCACT	TGTGCATAAC	4080
	CTAAGTCATG	TACAGCTCGA	GTAAGCGTTT	GTCCTTGCGT	ATCTAATACT	TGTGGTTGTA	4140
	ATGTGATATG	TAGTTCAATT	GTTTTCATTA	TTTAAATCC	TCCAATTTGT	TTAAAAATAT	4200
50	TTGATATGTT	TCAATCAGTG	ATCCAGTGTT	ATTTCTATAT	ACATCTTTAT	CAAAGTTTGC	4260
	ATTGGTAGCT	TTATCCCAAA	TTGACATGT	ATCTGGAGAT	ATTTTCATCCG	CTAACAAAAT	4320

55

	ATCCATTAAT TGTTCACA CATTATTAAT CTTTAATGCT TTGGATTTTA GTATTTCAAT	4440
	ATCTTCATCT GATGCTATAT TGAGCAATTT AACATGGTCA TCCGTTATCA ACGGATCATT	4500
5	TAACGCATCA TTTTATAGA AAAATTCTAC AAGTGGTTCT CTAAAACTT CACCATT TTC	4560
	AAAACCTAAA CGCTTTGTAA TAGATCCACT AGCAATATTA CGAACAACTA CTTCTAATGG	4620
	AATTATTTTC ACAGGCTTAA CTAATTGTTC TGTTCAGAT AATTGTTTAA TAAAGTGA	4680
10	TTCTATTCCA TTTCTTGTA AATATTTTAA TATAATAGAA GTAATTGAT TATTTAATCG	4740
	CCCCTTACCT GCCATTGTGT CTTCTTAGC CCGTTTCCA GCAGTAACTT CATCTTTATA	4800
	TTCAACTCTT AATTCATTTT CTGATTTGT TGAGAAAATG CGCTTCGCTT TTCCTTCATA	4860
15	TAATAATGTC ATGCTTTAAT TACTCCCCTC AAATTTAGCG TACATATCTT GTTCAGTTTG	4920
	GTTTACATCA TTCGTTAGTA CAGTCATATG CCCCATTTTT CTGCTATCTT TACGCTCAGA	4980
20	CTTACCATAA ATATGTAAGT GCCACTCTGG ATGTTCAATTA AATTCATTTT CCAATAAATC	5040
	TAAATCTTTA CCTAGTAAGT TCATCATGAC TGCTGGCTTT AATAATTCAA TTGAATTTGG	5100
	TAATGATTGT CCGGTAAGT CTAATAATG AGTATCAAAT TGTGAATAAT CACATGCTTC	5160
25	AATGAATAA TGTCCGGAAT TGTGAGGCCT TGGTGCTATC TCGTTCACAT ACAATTGGTT	5220
	GTTACTATCT ATAAAAAATT CAACTGTAAA TGTTCCAATG AAATGAATCG ATTGGATAAT	5280
	TTTATTAACT TGCTCTTTCG CCTCAGCTGT TTTATCTATT CTCGCTGGAA CAATTGTTTT	5340
30	GAAAAGTATT TGATTTCTAT GCTCATTTTC TTGTAATGGG AAAAAAGTGA TTTGATTGTT	5400
	GTTTCCTCTT GTAACAGTAA GAGATACTTC TTTCTTGATA TTCAAATATT TTCAGCTAC	5460
	GCATTCACCTA GTTTCAAITA ATTTAAAACC TTCTTGTAAG TCTTTTTTCGT TGTTAATTAA	5520
35	AACTTGACCT TTGCCATCGT AGCCACCAAA TCTAGTTTTT ACAATAAAAG GATATCCTAA	5580
	TGTTTCAATT GCTTTGTCAA TATCTGTAGA TTCTTTTACT GAAATGAACG GGACAACTTT	5640
	GGTACCAGCA CTTTTTAATG TTTCTTTTTC AGTTAAGCGA TCTTGTAATA ACTGTATAGC	5700
40	TTGGTAACCT TGCGGAATAT TGTACTTTTC ACATAATAGT TTTAATTGTT GGGCTGAAAT	5760
	GTTTTCAAAT TCATAAGTAA TCACATCACA TTTTGTGCTT AATTGATTGA GTGCCTTTTC	5820
45	ATCGTCATAC TTGGCTTGTA TAAATTCGTG TGCAACGTAT CTACATGGAC AATCTTCAGA	5880
	AGGATCCAAT ACAACCACTT TATAACCCAT TTTTGTGAGCT GATTGTGCCA TCATCTTTCC	5940
	AAGCTGACCA CCACCAATAA TGCCAATAGT CGCACCAAAC TTTAATTTAT TGAAGTTCAT	6000
50	TTTGCATGTC CTCCACTTTT TGAATTAACG AAGATTCATA CTGATTTAGT TTTTCAACTA	6060
	AAGAAGGATT TTGAATACTT AACATTCTTG CTGCAAGTAT ACCTGCGTTT TTAGCACCTG	6120

55

	AAGAATCTAT ACCCTTTAAA CTTTTTGTTT CAATCGGCAC TCCAATAACT GGTAGCGTCG	6240
	TTAATGATGC AACCATACCT GGTAATGTG CCGCACC GCC AGCGCCTGCA ATGATAATGT	6300
5	TTATACCTCT TTCTCTCGCT TCAGAAGCAA ATTGAACCAT CATTTTTGGC GTACGATGTG	6360
	CGGATACTAC TTGTTTTTCG TACGGAATTT CAAAATAATC CAACATGTTA CAACTCTCTT	6420
	GCATAATTTT CCAATCGGAA GAACTGCCCA TAATGACTGC TACTTTCCTT TTGTACACCC	6480
10	TTTCAAAAGT TTGAATTGTG AATTACTTTA GTTGTATATT ATAGATATAG CATAACAAGC	6540
	AATTTCTGCT TTTTCAATCA AAAATCGAAC TTTATTTTGA TTTTTTATTT GAATTTACGT	6600
	CTTTTGCTAT GTAAATTAGT TTTATAAACT AACAAAGTTA GGATATTGAC AATAGGAGGA	6660
15	GAAGTTTTTA TGGTTGCTAA AATTTTAGAT GGTAAACAAA TTGCCAAAGA CTACAGACAG	6720
	GGGTTACAAG ATCAAGTTGA AGCGCTAAAA GAAAAGGGTT TTACACCTAA ATTATCCGTT	6780
20	ATATTAGTTG GTAATGATGG CGCTAGTCAA AGTTATGTGA GATCAAAAAA GAAAGCAGCT	6840
	GAAAAAATTG GTATGATTTG AGAAATCGTA CATTTGGAAG AAACAGCTAC TGAAGAAGAA	6900
	GTATTAAACG AACTAAATAG ACTAAATAAT GATGATTCTG TAAGTGGTAT TTTGGTACAA	6960
25	GTACCATTAC CAAAACAAGT TAGCGAACAG AAAATATTAG AAGCAATCAA TCCTGAAAAA	7020
	GATGTGGACG GTTTTCATCC AATAAATATA GGGAAATTAT ATATCGATGA ACAAACTTTT	7080
	GTACCTTGCA CACCGCTCGG CATCATGGAA ATATTAAAC ATGCTGATAT TGATTTAGAA	7140
30	GGTAAAAATG CAGTTGTAAT TGGACGAAGT CATATTGTG GACAACCAGT TTCTAAGTTA	7200
	CTACTTCAAA AAAATGCATC AGTAACAATC TTACATTCTC GTTCAAAAGA TATGGCATCA	7260
	TATTTAAAAG ATGCTGATGT CATGTGTCAGT GCAGTTGGTA AGCCTGGTTT AGTAACAAAA	7320
35	GATGTGGTCA AAGAAGGAGC AGTAATTATC GATGTTGGCA ATACGCCAGA TGAAAATGGC	7380
	AAATTAAAAG GTGACGTTGA TTATGATGCG GTTAAAGAAA TTGCTGGAGC TATTACACCA	7440
	GTTCTTGGTG GCGTTGGTCC ATTAACAATT ACTATGGTAT TAAATAATAC TTTGCTTGCA	7500
40	GAAAAAATGC GTCGAGGTAT TGATTCGTAA AGAGCCTGAG ACATAAATCA ATGTTCTATG	7560
	CTCTACAAAG TTATAATGGC AGTAGTTGAC TGAACGAAAA TTCGCTTGTA ACAAGCTTTT	7620
45	TTCAATTCTA GTCAACCTTG CCGGGGTGGG ACGACGAAAT AAATTTTACG AAAATATCAT	7680
	TTCTGTCCCA CTCCCTAATA ACTGAGTTTT AATGAAGTCT TTTAACCAC ATTAAATATT	7740
	ATTTTGCAAT TGCAATGAAT AACAAGAAAA ATCTGGGACA TTAATCGATC AAATGCTCCC	7800
50	TTCAAAGTAG ACATTGAATA AATGAAGGCT TTGAAGGGAG CATTTCACTT TGTACTTGGC	7860
	TCAACAATTT TATATAGACA GTAGTTAATT GAATGAAAAT AAGCTTGTA CAAGTTTTCA	7920

55

GTTGGGGATG GGCCCCAACA CAGAAGCTGT GACTATGATA AAGTACTACT ACATAGTTAA 8040
 TCATTAGTGG TTCTTTATCA TTTTCGCCTC CCTTTTCTTA TTGTTTTGAT ACACAAAAAT 8100
 5 TTAAGTTCAA ACTGTCGAAT AAAGTTATAT TTGATTTCAA ATTATCCCTA AATTATTAAT 8160
 TKTACAATTG TGGCAGATTT TCAAAATAAT AATTATTTCC TCATTATTTA TAAATTTATA 8220
 TTTAAATTTT ATTCTTTATA GGGTAAGATT AGGACTATAG TATGATGTGT AATAATATA 8280
 10 AATTAAGGTA TAGTAAAGCT AACTCAGAAA TGACTTATCA TTCGGAGGTT ACATTATGAA 8340
 TAAACTATTA CAGTCATTAT CAGCCCTCGG TGTTTCTGCT AACTAGTAA CACCAAATTT 8400
 15 AAATGCAGAT GCAACGACGA ATACTACACC ACAAATTAAG GCGCTAATG ATATCGTTAT 8460
 TAAGAAAGGT CAAGATTATA ACCTTCTAAA CGGCATAAGT GCATTTGATA AAGAAGATGG 8520
 AGATTTAACC GATAAAATTA AAGTCGATGG CCAAATTGAT ACATCTAAAT CTGGTAAATA 8580
 20 TCAAATTAAG TATCATGTCA CTGATTGAGA TGGTGCAATT AAAATTTCCA CTAGGTATAT 8640
 TGAGGTTAA TAGCCCTCAT CACTATACTG CAAATAAAAT GGTAGCAAAC GAACATGTTT 8700
 TGCTACCATT TTATTTGTTA TTCTAACTTC ATCTGCAACT TTAACCCAAA TATTGTATTT 8760
 25 TTTCTGTATA CCAAAGGACT ACCTATCAAA TTATTAAAC TTAAGTCTC TTTTAAAAA 8820
 AATGTTTTGA TTTTGAACAA ACAAATTTCC ACTTTTCATT GTTTAACGAT AAATTACTTT 8880
 TGGCAAATTC CTTATTAAAA TGTTTGCCTC TCCTTTCAAT CAACTAGCCA TCATTTTCAA 8940
 30 TTTATTAGAC AATTTCAAAC TTTTTTTATT TTCATTCAAT TAACCTTTAA TTGAAAGCTA 9000
 TTCTCAACTT TCCTTTTAAA TATGAAGCAA TTTTTTCAAA AACGCTATTA GTCACAAAAT 9060
 GT 9062

(2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2738 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

AAATATTTTT TCAAACTAT GTGAAATGG aCCATGTCTa aATCATGTAA TAATGCAGyA 60
 CATAATGCCA ACGGTCTmTC TTTATTGTCC CATGCATCAT GACCAATAAA TGACTCATCA 120
 50 ATTAATCGTC TAACTATTTT ATACACACCT AAAGAATGTC CAAAGCGACT ATGTTCTGCT 180
 GTGTGAAAAG ATAGGTACAG TGTTCCTAGT TGTCTAATTC GACGTAACCT TTGGAATTCC 240

	TCTTTAAAAA CTTTTTCTTC TACTAATTTT AAATCTACAT ATGCGTTAGT CATTATTCCC	360
	CTCCTTTTTCG TTAAATATAA TATTTAATTT ACTTAAAAATG CTTTGTACAT AAGTGCTAAG	420
5	TCTAACTTTT CGCCATACAT TTCTGGCTCA TAAGAGCGTA AGATTGTAAA ACCTTGCTCT	480
	TTATAGTAAG CTACTGCTTC TTCATTTTTA TTATCTACTT CTAAGTAAAC ACCTTCAAAT	540
	TTATCTTCAA AACGTGATAA TCCTTCATTT AACAAATGCTG TACCATAACC TGTATGTTGC	600
10	GATTCTGGTT TAACATAATG AGCTGATAAA TATAATTCTT CACCGTAAAT AAAGTTAGCA	660
	AAGCCAACGA TGTCAATTACC TTCTTCAACG ACTAAGAATA ATTGTTCTTG AAGTCTTTTC	720
	TTTAAATGAT GTTCATTATA TGAAGCTtCT AACAAAGTGAT TAACTGTTGT CGCAGCGTAT	780
15	ATATTTAAGT ATGTATTAAA CCAAGCTTTA GTTGCGACAT CTCTAATTTG AACAAATCT	840
	TTTTCAAGTTG CTGTCTTAC CTTGAACATG ACTTTTCTCCC CTTATTAACA AGTTTAAATA	900
	ACGGCATTAT ACCACAACCTT GCTCAATACT TAATAAACAA TGATTGTCTA TTCAATTTAT	960
20	ATATtTATAT TTTCCGTTAA AATTAAAAAT AAAAAATAAC GAAGCAAAAA AtCACTTCGT	1020
	TTAGTATGAG GTATGTCTTA TTGCAATATA CTATTCCACT CAGTTGCACG TGCTAAGGCA	1080
25	TAGTTGTCTT TCATGATGTC ACCAGGCTTT TCAGCAGTTC CAATAATATA ACCATTTAAA	1140
	GTGGCACCTA rAAAGTCTAA ACTATATTTT ATTTGCGTAA TTGCTGGTTC GCTTTTATTT	1200
	TTGGACAATC TCCACCAACT AAAATAACTC TAAATCCTT TTCGGCCATT TGTGCCTTAA	1260
30	AATTAGGATA TCGTTTATCT TGTAATGTTT CTGACCAATG TTCGATAAAT GCTTTCAATG	1320
	GTGCTGAAAT GCTATACCAA TACACTGGTG ATGCAAAAAT AATTGTATCA CTAGCCAATA	1380
	TTTTATCTAG AATCGGCAAA TAGTCATCGT CATATGAAGT AATAGTCTCT GCTGTATGTC	1440
35	TCACGTCACG TATCGGTTTA AACTGATGTT GTGTCACGTC AATCCATTGA TACTCTAAAT	1500
	CTTGCAAAGC GAATTTTGTT AATTGTGCAG TATTACCGTT TGGTCTACTC CCACCAAACA	1560
	AAACAGTAAT CATTTTAGCC TAACCTCACT TTTGATTAAT AAATATCTGT GTTTTTCGTT	1620
40	ACCTAATTAT ACTATCATAA GCTTTGCCTA CCGAATAGTA AAACGCTTAC AACTTTTATA	1680
	TAAATTTGAC GAAATTTGCT CATGCCTTAT ATAACGTCGT TTGTGATACG GGGCTAATTC	1740
45	ATGATGAAAT TAGATACATA TATCACCATT AAATACAATT CATTTAGTCT TCAATCGGAA	1800
	ACAGTTCATC GATATATTGA ATCTCATCAT CTGATAAAAC GATATCTGCA GCTTTAATAT	1860
	TTTCAACGAC TTGTTCTGCA CGTTTTGCAC CAGGAATAAT CACATCGATA GCTGGTCTCG	1920
50	TTAAATAAAA TGCTAATACA ATGTTGCGAA TTGAAGTTTG ATGTGCTGCA GCTATGCTTT	1980
	CCAAAGCTTT TACGCGACGC ACATTTTCTT CAAATACACC TGGTTTAAAA TCACGACGTG	2040

55

GCTAATGGGA AATATGGAAT AAATGTGATT TGGTGATCAA CACAATATTG TAATACTGCC 2160
 TCATTTTCGC GATGCAATAA ATTATATTCT AACTGTACAA CATCAACGTA ACCATCTTTA 2220
 5 TTTGCTTCTT TAAGTTGATC TAATGTGAAA TTTGATACAC CAATTGCTTT AATCTTCCCT 2280
 TGTTCTTTAA GCTCTTGTA TGTGCAACT GCTTGATCTT TCGGAGTGTT GTTATCCGGA 2340
 AAATGAATAT AATATAAATC GATATAATCA GTTTGTAGAC GTTTCAAACT ATTCTCAACT 2400
 10 TGTTGTTTTA AATATTCCGG TTGATTGTTC TGATGTACTT CTTGATTTTC ATCAAATTCA 2460
 TGAGACCCTT TCGTAGCAAT TTTAATTGTC TCTCGCGGAT ATTCTTTAAC AACTTCTCCA 2520
 ACCAATTCTT CTGATCGTTC TGGCCCATAA ATATATGCCG TATCTAATAA ATTAATACCA 2580
 15 TGATTAATGG CTTGACGAAC AACATCTTTT CCTGTGTTCTT CATCTAAGTT CGGATATAAA 2640
 TTATGCCCAa CCTAtGCGTT CGTCCCAAGT GCGATTGGAA ACACTTCAAC ATCAGATTTA 2700
 20 CCTAAGTTTA CAAATTGCTn CATTAGACCC AGnCCTT 2738

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9425 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

GATTAGATGA TATTTAACGA AAATTaGrT GmAATACTtG AATGTArGaa GTCTGATGTC 60
 GAAAATAGCT ATTAAATAG AGTAGACGTA ATGtAAATGA AAGCACCTAA AATAGAAAAA 120
 35 TTTCAAAAAT AGCGTAATTA TTATAATAAA TAGACTGCCA ATAAAATGCA ATTTTTCACT 180
 TATAACATTC TTCAAAAAT AATAGCAAAA TTATGTAAAA AATATCTTGT CATGGCAAGA 240
 TTGGCTGTGC TATAATCTAT CTTGTGCTTA AGAACGGCTC CTTGGTCAAG CGGTTAAGAC 300
 40 ACCGCCCTTT CACGGCGGTA ACACGGGTTT GAGTCCCGTA GGAGTCACCA TTTTTTAGGT 360
 CTCGTAGTGT AGCGGTTAAC ACGCCTGCCT GTCACGCAGG AGATCGCGGG TTCGATTCCC 420
 GTCGAGACCG TACAAATGCC TATCCAAGAG GATAGGCATT TTTTTCGTT TAATATTATA 480
 45 TTAATAAAAAG ATATATGGAC GAATGATAAT CATATTGATT TATCTGTTCTG TCCATTTTCT 540
 TTAAAATGTA TGAACCTCAA GTAACCTAGT GGTGATAT GAAAGATAAA CGTAGACAAT 600
 50 AAAATCTTTA TTAGACGTAC AACATATGC TACTGTCAAC ATATTTCTTC GTTGTGATAT 660
 GCCACCAGTC CTCCATAACA TCAATTGTTA AAGTAACGAA TAACGAATAA TGATATTTAT 720

55

	GACCTCATCA TTGTGTTAAA TATCATTGTC ACAATCCGCC GTGAGAACT AATAAAAAAT	840
	AGTAATATAT AAGTTTATAT TGGAAAATAG AATTAATAGC TTATAAATGG TAAATTATAT	900
5	AATAGGTTAC TATACGTTAT AAGACGGAAA ATGCGCACAA TAACAAAAAT AGTAAGCGAC	960
	ATCCTGTGAT TTTTACACA AACATAAACG ATAAAGAACA AAAAATGATA AAATAATATT	1020
	AATGATTTAA GAAAAGAGGT TTATGCAAAT GGCTAGAAAA GTTGTGTAG TTGATGATGA	1080
10	AAAACCGATT GCTGATATTT TAGAATTTAA CTTAAAAAAA GAAGGATACG ATGTGTACTG	1140
	TGCATACGAT GGTAATGATG CAGTCGACTT AATTTATGAA GAAGAACCAG ACATCGTATT	1200
15	ACTAGATATC ATGTTACCTG GTCGTGATGG TATGGAAGTA TGTCGTGAAG TCGCAAAAA	1260
	ATACGAAATG CCAATAATAA TGCTTACTGC TAAAGATTCA GAAATTGATA AAGTGCTTGG	1320
	TTTAGAACTA GGTGCAGATG ACTATGTAAC GAAACCGTTT AGTACGCGTG AATTAATCGC	1380
20	ACGTGTGAAA GCGAACTTAC GTCGTCATTA CTCACAACCA GCACAAGACA CTGGAAATGT	1440
	AACGAATGAA ATCACAATTA AAGATATTGT GATTTATCCA GACGCATATT CTATTAAAAA	1500
	ACGTGGCGAA GATATTGAAT TAACACATCG TGAATTTGAA TTGTTCCATT ATTTATCAAA	1560
25	ACATATGGGA CAAGTAATGA CACGTGAACA TTTATTACAA ACAGTATGGG GCTATGATTA	1620
	CTTTGGCGAT GTACGTACGG TCGATGTAAC GATTCGTCGT TTACGTGAAA AGATTGAAGA	1680
	TGATCCGTCA CATCCTGAAT ATATTGTGAC GCGTAGAGGC GTTGGATATT TCCTCCAACA	1740
30	ACATGAGTAG AGGTCGAAAC GAATGAAGTG GCTAAAACAA CTACAATCCC TTCATACTAA	1800
	ATTTGTAATT GTTTATGTAT TACTGATTAT CATTTGGTATG CAAATTATCG GGTTATATTT	1860
	TACAAATAAC CTTGAAAAAG AGCTGCTTGA TAATTTTAAG AAGAATATTA CGCAGTACGC	1920
35	GAAACAATTA GAAATTAGTA TTGAAAAAGT ATATGACGAA AAGGGCTCCG TAAATGCACA	1980
	AAAAGATATT CAAAATTTAT TAAAGTGAAGTA TGCCAACCGT CAAGAAATTG GAGAAATTCG	2040
	TTTTATAGAT AAAGACCAAA TTATTATTGC GACGACGAAG CAGTCTAACC GTAGTCTAAT	2100
40	CAATCAAAAA GCGAATGATA GTTCTGTCCA AAAAGCACTA TCACTAGGAC AATCAAACGA	2160
	TCATTTAATT TAAAAGATT ATGGCGGTGG TAAGGACCGT GTCTGGGTAT ATAATATCCC	2220
45	AGTTAAAGTC GATAAAAAGG TAATTGGTAA TATTTATATC GAATCAAAAA TTAATGACGT	2280
	TTATAACCAA TTAAATAATA TAAATCAAAT ATTCATTGTT GGTACAGCTA TTTCAATTATT	2340
	AATgCACAGT CATCCTAGGA TTCTTTATAG CGCGAACGAT TACCAAACCA ATCACCATA	2400
50	TGCGTAACCA GACGGTCGAA ATGTCCaGAG GTAACATATAC GCAACGTGTG AAGATTTATG	2460
	GTAATGATGA AATTGGCGAA TTAGCTTTAG CATTTAATAA CTTGTCTAAA CGTGTACAAG	2520

55

	GTGATGGTAT TATTGCAACA GACCGCCGTG GACGTATTCTG TATCGTCAAT GATATGGCAC	2640
	TCAAGATGCT TGGTATGGCG AAAGAAGACA TCATCGGATA TTACATGTTA AGTGTATTAA	2700
5	GTCTTGAAGA TGAATTTAAA CTGGAAGAAA TTCAAGAGAA TAATGATAGT TTCTTATTAG	2760
	ATTTAAATGA AGAAGAAGGT CTAATCGCAC GTGTAACTT TAGTACGATT GTGCAGGAAA	2820
	CAGGATTGT AACTGGTTAT ATCGCTGTGT TACATGACGT AACTGAACAA CAACAAGTTG	2880
10	AACGTGAGCG TCGTGAATTT GTTGCCAATG TATCAGATGA GTTACGTACA CCTTTAACTT	2940
	CTATGAATAG TTACATTGAA GCACTTGAAG AAGGTGCATG GAAAGATGAG GAACTTGCGC	3000
	CACAAATTTT ATCTGTTACC CGTGAAGAAA CAGAACGAAT GATTGACTG GTCAATGACT	3060
15	TGCTACAGTT ATCTAAAATG GATAATGAGT CTGATCAAAT CAACAAAGAA ATTATCGACT	3120
	TTAACATGTT CATTAATAAA ATTATTAATC GACATGAAAT GTCTGCGAAA GATACAACAT	3180
20	TTATTCGAGA TATTCGAAA AAGACGATTT TCACAGAAAT TGATCCTGAT AAAATGACGC	3240
	AAGTATTTGA TAATGTCATT ACAAATGCGA TGAAATATTC TAGAGGCGAT AAACGTGTCG	3300
	AGTTCACGT GAAACAAAT CCACTTTATA ATCGAATGAC GATTGCTATT AAAGATAATG	3360
25	GCATTGGTAT TCCTATCAAT AAAGTCGATA AGATATTCGA CCGATTCTAT CGTGTAGATA	3420
	AGGCACGTAC GCGTAAAATG GGTGGTACTG GATTAGGACT AGCCATTTCG AAAGAGATTG	3480
	TGGAAGCGCA CAATGGTCGT ATTTGGGCAA ACAGTGTAGA AGGTCAAGGT ACATCTATCT	3540
30	TTATCACACT TCCATGTGAA GTCATTGAAG ACGGTGATTG GGATGAATAA TAAGGAGCAT	3600
	ATTAAATCTG TCATTTTAGC ACTACTCGTC TTGATGAGTG TCGTATTGAC ATATATGGTA	3660
	TGGAACTTTT CTCCTGATAT TGCAAATGTC GACAATACAG ATAGTAAGAA GAGTGAAACG	3720
35	TAACCTTTAA CGACACCTAT GACAGCCAAA ATGGATACAA CTATTACGCC ATTCAGATT	3780
	ATTCATTCTGA AAAATGATCA TCCAGAAGGA ACGATTGCGA CGGTATCTAA TGTGAATAAA	3840
	CTGACGAAAC CTTTGAAAAA TAAAGAAGTG AAGTCCGTGG AACATGTTCTG TCGTGATCAT	3900
40	AACITGATGA TTCCTGATTT GAACAGTGAT TTTATATTAT TCGATTTTAC GTATGATTTA	3960
	CCGTTATCAA CATATCTTGG TCAAGTACTG AACATGAATG CGAAAGTACC AAATCATTTT	4020
45	AATTTCAATC GTTTGGTCAT AGATCATGAT GCTGATGATA ATATCGTGCT TTATGCTATA	4080
	AGCAAAGATC GCCACGATTA CGTAAAATTA ACAACTACAA CGAAAAATGA TCATTTTTTA	4140
	GATGCATTAG CAGCAGTGAA AAAAGATATG CAACCATACA CAGATATCAT CACAAACAAA	4200
50	GATACAATTG ATCGTACGAC GCATGTTTTT GCACCAAGTA AACCTGAAAA GTTAAAAACA	4260
	TATCGCATGG TATTTAACAC GATTAGTGTT GAGAAAATGA ATGCTATACT ATTTGACGAT	4320

55

	GCAAACATA ACGATAAAAA TGAAAAATAT CATTATAAAA ACCTGTCCGA AGATGAAGCG	4440
	AGTTCACGCA AAATGGAAGA AACGATTCCA GGAACCTTTG ATTTTATTAA TGGTCATGGT	4500
5	GGTTTCTTAA ACGAAGACTT TAGATTGTTT AGTACGAATA ATCAGTCAGG CGAGTTAACA	4560
	TATCaACGTT TCctTAATGG TTATCCAACG TTTAATAAAG AAGGTTCTAA TCAAATTCAA	4620
	GTCACCTGGG GTGAAAAAGG CGTCTTTGAC TATCGTCGTT CGTTATTACG CACCGACGTT	4680
10	GTTTTAAATA GTGAGGATAA TAAATCGTTG CCGAAATTAG AGTCTGTACG TTCAAGCTTA	4740
	GCGAACATA GTGATATTAA TTTTGAAAAA GTAACAAACA TCGCTATCGG TTACGAAATG	4800
15	CAGGATAATT CAGATCATAA TCACATTGAA GTGCAGATTA ACAGTGAACCT CGTACCGCGT	4860
	TGGTATGTAG AATATGATGG CGAATGGTAT GTTTATAACG ATGGGaGGCT TGaATAAATG	4920
	AACTGGaAAC TGACAAAGAC ACTTTTCATT TTCGTGTTTA TTCTGTCAA CATCGTGTTA	4980
20	GTATCGATTT ATGTTAATAA AGTCAATCGC TCACACATTA ATGAAGTCGA GAGTAACAAT	5040
	GAAGTTAATT TTCAGCAAGA AGAAATTAAA GTACCGACTA GTATATTGAA TAAATCAGTT	5100
	AAAGGTATAA AATTAGAGCA AATTACAGGG CGATCAAAAG ACTTTAGTTC TAAAGCTAAA	5160
25	GGCGATTCGG ATTTGACCAC ATCAGATGGT GGAAAATTAT TGAATGCGAA CATTAGTCAA	5220
	TCGGTAAAGG TCAGTGACAA TAACTTAAAA GATTTGAAAG ATTATGTTAA CAAGCGCGTA	5280
	TTTAAAGGTG CTGAATATCA ATTAAGCGAG ATTAGTTCAG ATTCTGTAAA ATATGAACAA	5340
30	ACGTATGATG ATTTTCCGAT TTTAATAAAC AGTAAAGCGA TGTAAACTT TAATATAGAA	5400
	GATAACAAAG CGACTAGTTA TAAACAATCA ATGATGGATG ACATTAAGCC CACAGATGGT	5460
	GCAGATAAGA AGCATCAAGT GATTGGTGTG AGAAAAGCAA TCGAGGCATT ATATTATAAT	5520
35	CGTTACTTGA AAAAAGGTGA TGAAGTCATT AATGCTAGAC TCGGTTACTA CTCAGTCGTG	5580
	AATGAAACGA ATGTTCAATT GTTACAACCA AACTGGGAAA TTAAAGTGAA GCATGACGGT	5640
40	AAGGATAAAA CGAATACTTA CTATGTCGAA GCGACAAATA ATAACCCTAA AATTATTAAT	5700
	CATTAATATG AATCGTAATA AGCTAGCATT GCAAGTCAT CATATGTGAG AAGCGGTGCT	5760
	AGCTTTTTTG CTGGTACGGT TTATTATGGC TGATGTTTTT GCGTCTCAA CGTGCGCATT	5820
45	TATTCATATT TTAAGTAGAA CCGCATGTGA AAATTAGTGT AACTGTTATT TTAAAACTT	5880
	TAGTATTTGT CTAATCATTG TTATAATAAT TAAGAAATTC ATTGCACGTG ATTATCAAAA	5940
	TTTAAATATA AGAAACCGGT CGATGAACTA AAGTTACATA ATAGGAAAGG TATACAAAAC	6000
50	AGCTAATATA CTGATAGTTT CTGTAGGGAA AATCGTATAT TTGCACTGAT GTATATTGCA	6060
	GTCATATAGA GAGATTGACT GTTTAAAGAG AAAGGATGAG CCGCTTGATA CGCATGAGTG	6120

55

	TAGTTGATGT TGGTTTGA	CT GGAAAGAAAA	TGGAAGAATT	GTTTAGTCAA	ATTGACCGTA	6240
	ATATTCAAGA TTAAATGGT	ATTTTAGTAA	CCCATGAACA	TATTGATCAT	ATTAAAGGAT	6300
5	TAGGTGTTTT GCGCGTAA	TATCAATTGC	CAATTTATGC	GAATGAAAAA	ACTTGGCAGG	6360
	CAATTGAAAA GAAAGATAGT	CGCATCCCTA	TGGATCAGAA	ATTCATTTTT	AATCCTTATG	6420
	AAACAAAATC TATTGCAGGT	TTCGATGTTG	AATCGTTTAA	CGTGTCACAT	GATGCAATAG	6480
10	ATCCGCAATT TTATATTTTC	CATAATAACT	ATAAGAAGTT	TACGATTTTA	ACGGATACGG	6540
	GTTACGTGTC TGATCGTATG	AAAGGTATGA	TACGTGGCAG	CGATGCGTTT	ATTTTGTAGA	6600
15	GTAATCATGA CGTCGATATG	TTGAGAATGT	GTCGTTATCC	ATGGAAGACG	AAACAACGTA	6660
	TTTTAGGCGA TATGGGTCAT	GATCTAATG	AGGATGCGGC	TCATGCAATG	ACAGACGTGA	6720
	TTACAGGTAA CACGAAACGT	ATTTACCTAT	CGCATTTATC	ACAAGACAAT	AACATGAAAG	6780
20	ATTTGGCGCG TATGAGTGTT	GGCCAAGTAT	TGAACGAACA	CGATATTGAT	ACGGAAAAAG	6840
	AAGTATTGCT ATGTGATACG	GATAAAGCTA	TTCCAACGCC	AATATATACA	ATATAAATGA	6900
	GAGTCATCCG ATAAAGTTCC	GCATTGCTGT	GAGACGACTT	TATCGGGTGC	TTTTTTATGT	6960
25	TGTTGGTGGG AAATGGCTGT	TGTTGAGTTG	AATCGGCTTG	ATTGAAATGT	GTAAATAAT	7020
	TCGATATTAA ATGTAATTTA	TAAATAATTT	ACATAAAATC	AATCATTTTA	ATATAAGGAT	7080
	TATGATAATA TATTGGTGTA	TGACAGTTAA	TGGAGGGAAC	GAAATGAAAG	CTTTATTACT	7140
30	TAAACAAGT GTATGGCTCG	TTTTGCTTTT	TAGTGTAATG	GGATTATGGC	AAGTCTCGAA	7200
	CGCGGCTGAG CAGCATACAC	CAATGAAAGC	ACATGCAGTA	ACAACGATAG	ACAAAGCAAC	7260
	AACAGATAAG CAACAAGTAC	CGCCAACAAA	GGAAGCGGCT	CATCATTCTG	GCAAAGAAGC	7320
35	GGCAACCAAC GTATCAGCAT	CAGCGCAGGG	AACAGCTGAT	GATACAAACA	GCAAAGTAAC	7380
	ATCGAACGCA CCATCTAACA	AACCATCTAC	AGTAGTTTCA	ACAAAAGTAA	ACGAAACACG	7440
40	CGACGTAGAT ACACAACAAG	CCTCAACACA	AAAACCAACT	CACACAGCAA	CGTTCAAATT	7500
	ATCAAATGCT AAAACAGCAT	CACTTTCACC	ACGAATGTTT	GCTGCTAATG	CACCACAAAC	7560
	AACAACACAT AAAATATTAC	ATACAAATGA	TATCCATGGC	CGACTAGCCG	AAGAAAAAGG	7620
45	GCGTGTATC GGTATGGCTA	AATTAAAAAC	AGTAAAAGAA	CAAGAAAAGC	CTGATTTAAT	7680
	GTTAGACGCA GGAGACGCCT	TCCAAGGTTT	ACCACTTTCA	AACCAGTCTA	AAGGTGAAGA	7740
	AATGGCTAAA GCAATGAATG	CAGTAGGTIA	TGATGCTATG	GCAGTCGGTA	ACCATGAATT	7800
50	TGACTTTGGA TACGATCAGT	TGAAAAAGTT	AGAGGGTATG	TTAGACTTCC	CGATGCTAAG	7860
	TACTAACGTT TATAAAGATG	GAAAACGCGC	GTTTAAAGCCT	TCAACGATTG	TAACAAAAAA	7920

55

	TGAAGGCATT AAAGGCGTTG AATTTAGAGA TCCATTACAA AGTGTGACAG CGGAAATGAT	8040
	GCGTATTTAT AAAGACGTAG ATACATTTGT TGTTATATCA CATTTAGGAA TTGATCCTTC	8100
5	AACACAAGAA ACATGGCGTG GTGATTACTT AGTGAAACAA TTAAGTCAAA ATCCACAATT	8160
	GAAGAAACGT ATTACAGTTA TTGATGGTCA TTCACATACA GTACTTCAAA ATGGTCAAAT	8220
	TTATAACAAT GATGCATTGG CACAAACAGG TACAGCACTT GCGAATATCG GTAAGATTAC	8280
10	ATTTAATTAT CGCAATGGAG AGGTATCGAA TATTAAACCG TCATTGATTA ATGTTAAAGA	8340
	CGTTGAAAAT GTAACACCGA ACAAAGCATT AGCTGAACAA ATTAATCAAG CTGATCAAAC	8400
	ATTTAGAGCA CAAACTGCAG AGGTAATTAT TCCAAACAAT ACCATTGATT TCAAAGGAGA	8460
15	AAGAGATGAC GTTAGAACGC GTGAAACAAA TTTAGGAAAC GCGATTGCAG ATGCTATGGA	8520
	AGCGTATGGC GTTAAGAATT TCTCTAAAAA GACTGACTTT GCCGTGACAA ATGGTGGAGG	8580
20	TATTCGTGCC TCTATCGCAA AAGGTAAGGT GACACGCTAT GATTTAATCT CAGTATTACC	8640
	ATTTGGAAAT ACGATTGCGC AAATTGATGT AAAAGGTTCA GACGTCTGGA CGGCTTTCGA	8700
	ACATAGTTTA GGCGCACCAA CAACACAAAA GGACGGTAAG ACAGTGTTAA CAGCGAATGG	8760
25	CGGTTTACTA CATATCTCTG ATTCAATCCG TGTTTACTAT GATATAAATA AACCGTCTGG	8820
	CAAACGAATT AATGCTATTG AAATTTTAAA TAAAGAGACA GGTAAGTTTG AAAATATTGA	8880
	TTTAAAACGT GTATATCACG TAACGATGAA TGACTTCACA GCATCAGGTG GCGACGGATA	8940
30	TAGTATGTTT GGTGGTCCTA GAGAAGAAGG TATTTTATTA GATCAAGTAC TAGCAAGTTA	9000
	TTTAAAAACA GCTAACTTAG CTAAGTATGA TACGACAGAA CCACAACGTA TGTATTAGG	9060
	TAAACCAGCA GTAAGTGAAC AACCAGCTAA AGGACAACAA GGTAAGCAAAG GTAGTAAGTC	9120
35	TGGTAAAGAT ACACAACCAA TTGGTGACGA CAAAGTGATG GATCCAGCGA AAAAACCAGC	9180
	TCCAGGTAAA GTTGTATTGT TgtAGCGCAT AGAGGAACTG TTAGTAGCGG TACAGAAGGT	9240
40	TCTGGTCGCA CAATAGAAGG AGCTACTGTA TCAAGCAAGA GTGGGAAACA ATTGGCTAGA	9300
	ATGTCAGTGC CTAAAGGTAG CGCGCATGAG AAACAGTTAT TTCATAATCA ACAGTCATTG	9360
	ACGTAGCTAA GTAATGATAA ATAATCATAA ATAAAATTAC AGATATTGAC AAAAAATAGT	9420
45	AAATA	9425

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3886 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

	AGTTGTAATG TCACATTTCC AGAGTCTGAA ATTATCTTTA TCACGTTACA TTTACTAGGC	60
5	TCTAAAATGA CTGAACATAC AGCATCTTCA ATTACCTTTG AATACCATGA TTTATCGCAA	120
	AATATACATG AATTGATCAC TTGTGTTAGC CAAGAATTAG GCATTGATAT GTCAAAAGAC	180
	AACAAGTTAC ATACCAGTCT GATCACACAT ATCAAACCAG CTATACATCG TATTAAATAC	240
10	GATATGCTAC AACCTAATCC TTTGAGGCAA GAAGTTATGC GTCGCTATCC TCAAATCATT	300
	GAAGCCGTTA GCAAGCATAT TAGTCCAATT GAACAAGATG CTGCTATTCTG CTTCAACGAA	360
	GATGAATTAA CATACATTAC AATTCACCTC GCATCAAGTA TAGAGCGTGT TGCAACACAT	420
15	AAACAATCAA TGATTAAGGT TGTCTTACTA TGTGGTTCTG GTATAGGCAC GTCACAACCTT	480
	TTAAAATCAA AACTAAATCA CCTGTATCCT GaGTTnCACa TTTGGGAtGc CTATTcCATT	540
20	TaTcAATTGG aAGaAAGTCG ATTATTGCAA GATAACATTG ATTATGTCAT TTCAACAGTA	600
	CCTTGTAATA TATCAGCTGT ACCAGTTATT CATGTCGATC CATTTATCAA TCAACAATCT	660
	CGTCAAAAAT TGAATCAAAT TATCAATGAC TCAAGAGAAC AACGAGTCAT GAAAATGGCA	720
25	ACTGATGGCA AGTCACTCGC AGATTTATTG CCTGAACATC GCATCATTAT AAATAAACAA	780
	CCATTATCAA TTGAATCCGC AATTGCAGTG GCTGTGCAAC CTTTAATCAA TGATGGCATT	840
	GTCTATTCAA ATTATACAGC TGCAATTTTA AAACAATTG AACAAATTCGG GTCATATATG	900
30	GTCATTAGTC CACATATTGC ACTTATTCAC GCTGGTACTG ATTATGTACA GAATGGTGTA	960
	GGTTTCGCAC TAACATATTT CACTGAAGGG ATTATCTTTG GTAGTAAAGC TAACGATCCC	1020
	GTTACCTTG TAATTACATT AGCAACGGAC CACCCCAATG CACATTTAAA GGCATTGGGA	1080
35	CAGTTAAGCG AATGCTTAAG CAACGACTTA TATCGACAAG ATTTCTTAGA TGGGAATATT	1140
	TTTAAATTA AACAAACACAT TGCTTTAACT ATGACAAAGG AGGCTTAATA ACGTGTCAAT	1200
	AGACATTTTG TCAACAACAC GCATCATTGT AAAAGAACAA GTAAATGATT GGAATGAAGC	1260
40	TATAACTATA GCTTCTCAGC CATTACTACA AGAACAAATT ATTGAACAAG GCTATGTTCA	1320
	AGCAATGATT GATAGCGTTA ATGAACCTGG ACCTTATATC GTTATCGCAC CTGAAATTGC	1380
45	AATTGCACAT GCAAGACCGA ACAATGACGT ACATCAAGTT GGTTTAAGTC TATTAAAGTT	1440
	GAATCAACAT GTGGCATTCT GTGATGAAGA TCACTACGCA TCTCTCATTT TTGTATTGAG	1500
	TGCCATCGAC AATCATTCAC ACTTATCTGT ATTACAAAAT TTAGCAACCG TACTGGGCGA	1560
50	TAACCAAACA GTCCAGCAAC TATTAACTGC AACAAATGCA CAAGACATTA AAAACATTTT	1620
	AAAGGAGCAT GATTAATATG AAAATTTTAG TAGTATGTGG CCACGGTTTA GGAAGTAGTT	1680

55

	AAGTTGAACA TAGTGACATT ATGACAGCAA GTCCAGAGAT GGCTGACTTG TTTATTTGTG	1800
	GTAGAGATTT AGCTGAAAAT GCCGAACGTC TAGGGGATGT CTTAGTTCTT GATAATATTT	1860
5	TAGATAAAGC TGAATTACAA CAAAAGCTCT CAGAAAAATT ACAACAACCTT AACATGATTT	1920
	AAAGGAGGTA CGACCTATGC AAGCAATCCT TAATTTTATA GTCGATATTT TAAGTCAACC	1980
10	AGCCATTCTT GTTGCACTGA TTGCCTTTAT AGGTTTAAATC GTTCAGAAAA AACCTGCCGC	2040
	AACGATCACT TCAGGAACCA TTAAAACGAT ATTAGGCTTC TTAATTTTAA GTGCAGGTGC	2100
	TGATGTCGTC GTTCGATCTC TTGAACCATT CGGCAAAATA TTCCAACACG CATTTGGTGT	2160
15	GCAAGGTATC GTACCTAACA ACGAAGCTAT CGTCTCACTA GCCTTAAAAG ATTTTGGAAC	2220
	AACAGCTGCA CTCATCATGG TCTGTGGCAT GATTGTTAAT ATTTTAATTG CCCGCTTCAC	2280
	TAATTTAAAA TATATCTTTT TAACAGGTCA TCATACATTT TACATGGCTG CGTTTTTAGC	2340
20	AATCATTTTA ACAGTCAGTC ATATTAAAGG CTGGCTAACG ATTGTTATCG GCGCACTCGT	2400
	ATTAGGATTA ATCATGGCAG TATTACCTGC ATTACTCCA CCTACGATGC GAAAAATTAC	2460
	AGGGAATGAC CAAGTAGCTT TAGGTCATTT TGGCTCAATC AGTTACTTTG CCGCAGTGCT	2520
25	GTAGGTCAAT TATTCAAAGG TAAGTCTAAA TCAACGGAAG AGATTAAATT TCCAAAAGGC	2580
	TTAAGTTTCT TACGAGAAAG TACAATTAGT ATCTCGATTA CGATGGCATT ACTTTACTTC	2640
	ATCGCATGCT TATTTGCGGG CGTTAGTTAT GTACACGAAT CTATTAGTGA TGGTCAAAAC	2700
30	TTTATTGTCT TTTCATTAAT TCAAGGTGTG ACATTTGCTG CTGGTGTATT TATTATTTTA	2760
	ACGGGCGTTC GTTTAATCTT AGCTGAAATC GTCCCAGCAT TTAAAGGAAT TTCTGAAAAG	2820
35	CTTGTAACAA ATTCTAAACC TGCATTAGAC TGCCCTATTG TGTCCCTTA TGCACAAAAT	2880
	GCAGTATTAA TTGGATTCTT TGTCAGCTTT ATTACAGGTG TCATCGGTAT GTTTATCTTA	2940
	TTCTTATTTG GTGGCGTCGT CATTTTACCT GCGTAGTTG CACACTTCTT CTTAGGTGCA	3000
40	ACGGCTGCTG TATTCGGTAA TGCAAGAGGC GGTATTAAAG GTGCTATTGc TGGCGCCGCT	3060
	CTAAATGGTA TCCTAATCAC GTTTTTACCA TTATTATTCT TGCCATTTT AGGCGAATTA	3120
	GGTGGTGCTG CAACAACATT CTCAGATACA GACTTTTATG CTGTCGGTAT CGTGTTCCGGT	3180
45	AACGCAGTAA AATATATGGG ATTATTTGGT GCGATTCTAT TTATTATTAT CGTAGGTGCG	3240
	ACAACAATTT TATTAAAAGG CCGTCAAAAA GAACAGCAAT AGTGTTAACG TAGAAATATA	3300
	AAACACCGTC ACATATTGAG TGAATGCCCC TTTtATCAAG AGGAAAGCCA CTTACTTATG	3360
50	GACGGTGT TT TGTATTATAT TAAATGATAC TTAGCCATAC TATCGACAGC TGCTAAAATT	3420
	GCTTCTTCTT GTGTCGCAAT CGGTTCCCAA CCAAGTAATG TTTTgCACG TTCGTTACTT	3480

55

CCTAGACTCA AAATAAAGTC TGGTAATTTT TTAGTAGAAA CTTTTTGAGC TATTTTCAGGT 3600
 CTCTTTTCTT TAATTAATTT TGCAATTTCC AACAAATTAA TTTGTCCATC AGCCGTCGCA 3660
 5 ATAAATCGCT TGCCATTAGC TTGTTCAATT GTCAATTGCCA AAATGTGCAG TTCAGCTACG 3720
 TCTCTCACAT CAACAACATT TAACGGAATT TGCAGTACAC GTTTCATTGA ACCATTCAAT 3780
 AAATTTTCTA ATAAATGAAA GCTTCCTGAA ACGTGTGCAT CTAATGATGG CCCAAAAATT 3840
 10 GCAACTGGAT TGATTGTGGC AAATTCTACT GTTGTATTTT CATTCT 3886

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4879 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GTCACTCTATC AAAAATTTGG TATACAGACC GACAATTATT AATTAATAAT TTAATTTCCC 60
 25 AGGCAATACC AGTGATTAAA TATCCACAAA TACAACATAA AGAACAACCA TTAGAATCTA 120
 TTTCACAACT TATATTGTCT AAGATGACAT CTAATCAATA GTGTTTAAAT TTCTCAGTGG 180
 CTGTGAATGA GGTTTAAAG TACTATAAAA CGTAACTTT GATACTTTAA AATACGCAAA 240
 30 AAACGGTAAA CCCTAATTCA TATTATAGAG TTTACCGTTT TATTTTTTAA CTTGCATCAT 300
 AGTTATATTA ACATTATTGT TGGTAGTTTG GATCAGTAAC CATGCTTGT CCAGTATAAT 360
 CAACCGTTAC AATTGAATAT TTTCCATTG CATTGGGTC TTTAAACTA AACACATACT 420
 35 TATAGTTGCC ATTATGTTCT TCAATAGAAT AATCATTATA CACTTTATTA TTACTACCAA 480
 ATTTATTTGC TTCATTATTA GCCGCATTTA AAGCTGTTTG GAAATTTGGC AATTGCTGTA 540
 AAGCTTGATT TTTATTTCCA TTAAACGGAT AAATTGACG TGCAACCGGC GCGGCATTTT 600
 40 GNCATAATA TGGTGCAACG TAACCTGATT TTTGATTATT ATTCGCTTGG TTATTACTTG 660
 ATTGTTTATT ATTTGTTTGG TTTTGGTCAT TGTTGTTGC ATTTGAATTA GATTGTTGCT 720
 GGTATCGTT TGCATATTA TCTTTATTAT CTTTGTTTAC GTCTTTACTA TCATCTTTAT 780
 45 TATCTTTCTT ATCTTTAGAT GAATCATTG TTTTTTATC TTGTTGTTCA GTTTTCGCTT 840
 TATCATCTTT TTCTTTATTA CCGTCTTTT GTTGGTCACT ATCTTGACCA CATGCAGCTA 900
 50 AAAATAATGA TAATGCTAGT AACCCTGTAA CTAATCTTTT CATACATATC TCCTCCTATA 960
 ATTCGATATT CATTGAATAA TCTTGAAATA CATATCTACC ATGTGTATCT TTTCATGGCT 1020

	TAAGGTCTT TTTATTATAC CCTAATTTTT GTTCATTATT ATTTAATTTT TGTGAATTTT	1140
	ATGtTTkCTA TAAATTTAAT TATTTTACTT TAACAATTCA TTACGCATTT AGCATTTCAA	1200
5	GGTATACACA ATATTTATTA CTATGATTTT ATTTTATCTG CTGCAAAAAC AATCATTATA	1260
	ACTCTTTTTC CATAATTAAA TCTGTATCCG TTACATCACC TGTTTGAAAA TGATGTTTAC	1320
	CAACCACTTT AAATCCATGA CGTTTATAAA ATGCTTGAGC ACGAGGATTA TGCTCCCAA	1380
10	CTCCTAGCCA AATTTTATGT TTATTATGTT CTTGAGCAAT TTTTTCGGCC AATTCTATCA	1440
	ATTGTGAACC TCTTCCGCCA CCTTGAAAGT CTTTCAAAAA ATATATGCGC TGCATTCTA	1500
15	AATAGGTCTC CCCCATTTCT TCAGTTTGAG CACTATTAAT ATTCATCTTT ATATAACCAA	1560
	CATTGCGACC ATCTTCTTGa TAAAAATAAT GAAATGAATC TACATGGTTA ATCTCTGTG	1620
	TAAATTTCTC TACAGTATAA TTGTCTTTAA AAAATTGATC AAAATCTTTG TCATCATAGT	1680
20	AAGAACCAAA CGTGCATAA AATGTTCTAG TTGCTAATTC AACTAATTCA CTAGCATTTT	1740
	GTTCTGAAAT TTCTTTGATT ATCCCAGCCA TATAAATCCT CCAATAAACA GTGATCGAAT	1800
	CAAAATATTA CTTATGTTAT TTTTCAGCCA AAATATTTA AAAATACATT AACACAAATC	1860
25	AATTACAAAT TGTATTGATT GTGTGTAACA TCAATAAATG ATACATTTAT TCCAGTAAAA	1920
	TGGCCGTATT TTCAAAGAG AAAAAGAGAG GATGTATCGT TGTGATAGAA ACATTTAAAG	1980
	CGTTTGTAAT TGATAAAGAT GAGAGTGGTA AAGTGACACC AACTTTCAA CAATTATCGC	2040
30	CTACTGATTT ACCTAAAGGA GATGTGCTGA TTAAAGTACA TTAATCTGGT ATAAATTATA	2100
	AAGATGCTTT AGCGACTCAA GATCATAATG CAGTCGTAAT ATCGTATCCT ATGATTCCAG	2160
	GAATAGATTT AGCTGGAACA ATTGTTGAAT cCGAAGCACC AGGCTTTGaa AAAGGAGAAC	2220
35	AAGTAATTGT AACGAGTTAT GACCTAGGTG TCAGCCATTA TGGCGGTTTT AGTGAATATG	2280
	CGCGTGTAAT ATCAGAATGG ATTATCAAGC TTCCTGATAC TTTAACATTA GAAGAATCAA	2340
40	TGATATATGG CACAGCTGGT TATACTGCCG GTTTAGCAAT TGAAAGACTT GAAAAAGTTG	2400
	GAATGAATAT TGAAGATGGT CCTGTACTCG TTCGCGGTGC TTCAGGTGGT GTCGGTACTT	2460
	TAGCAGTACT CATGCTTAAT GAACTTGGTT ATAAAGTTAT CGCAAGTACA GGTAAACAAG	2520
45	ATGTTAGCGA TCAATTACTT GAACTTGGTG CCAAAGAAAT TATCGATCGA CTCCTGTTG	2580
	AAGATGATCA TAAAAAGCCA CTCGCATCAT CAACTTGGCA AGCTTGTGTA GACCCTGTTG	2640
	GTGGCGAAGG TATTAATTAT GTTACAAAGC GTTTAAATCA TAGTGGGTCA ATTACAGTTA	2700
50	TTGGTATGAC TGCCGGTAAT ACTTATACTA ATTCTGTATT CCCTCACATT TTAAGAGGTG	2760
	TAAACATTTT AGGAATTGAC TCGGTATTTA CTGCTATGAA ATTAAGACAG CGCGTTTGGC	2820

55

	TTGATGAACT TCCAGAACAA CTTAACAAAG TAATTAAACA TGAAAATAAA GGGCGCATTG	2940
	TTATCGATTT CCGTGTAGAT AAATAGTATT CATGAAAAAG ACATCCCGTT ATGCGAGATG	3000
5	TCTTTTTTAA TTTAGTATTT GATATACATA CCGCCTGAAT CTGGTTCGGT AGGTATAAAT	3060
	CCAAATTTTG TATATAATTT ATCCGCTGGG TAGTCTGCAA TCAGAcTAAC GTATGTACTC	3120
	TCAACAGCCA CACCTTTAAT ATATTGCATA ATATGCTCCA TAATTAGACT GCCGTAACCT	3180
10	TGACCTTGGT AACTTTTCAA AACTGCAATA TCAACAATTT GAAAAACAGT TCCGCCATCG	3240
	CCAATCACTC TACCCATACC AATTAACCGA TCTTTATCAT ACAAGGTAC TGTAATAAG	3300
	GCATTAGGTA ATCCTTTTTC aGCTGTTGCG GCGTCTTGG ACTCATACCT GCGTTAATCC	3360
15	TTAATGCGCA ATAATCCTCG CAAGTCGGAA TATCATATGT CACTTTAACC ATTATTTACC	3420
	CCACTTTTCA TCACACAATA TATCAACCTA GTATAAATGT TTATTTACAA TAGTCTTATT	3480
20	CGCTTCTTTA AACACTTCAT GATGACTTGA AACATAACCC TCTGCATTCTG CATCTGGTTG	3540
	GATATATGTT TTAGCAAGGT TCGCTGCATT TGCACCATCA CTAAATGCAC TTGCAATTAG	3600
	ATGTGATTTT GCATCATGAT AAACAATATC TCCACACGCA TAGATAACAG GTATACTAGT	3660
25	TGTCGTATTa CCAAATCCTT TAACACGACA ATCATCATGC ATATCTAGCT TTGAAGATGT	3720
	TtCACTCAAT AATGTATTAC AACGATCAAA CCCATGACTA ATAATGACAT CGTCAAATTT	3780
	AACTGTATGC CTATCGCCAC TTTCAACATG TTCCAAAACA ACTTCACTTA TATGCGTTTC	3840
30	ATCATCATTG CCGACCAAGT ATTTAATACG TGTTTTTGGG CATAGTTTCA CATTTAAATC	3900
	TGTCACCAAC GTTTTCATCG CTTTCATGACC ACTTACATCT TCTTTTCGAT AAACAACCTGT	3960
	CACGCTTTTA GCAATCTTGG CAATATCATG CGCCCAATCT AATGCTGTAT TTCCTCCACC	4020
35	TGATATTAAT ACATCTTTAT CTTTGAAACG TCTGTAACCT TGTAACAACAT AATGTAAATT	4080
	AGTTAATTGA TATCTCTCTA CACCTTTAAC ATCTAATTGT TTTGGATTAA TAATACCCGC	4140
	ACCAATTGCA ATGATAACTG CTTTCGATGT ATATATTTCT CCGCTTCTG TTTCAACTTC	4200
40	GAAATGACGT TCTGCCTTTT TCCTAATATC TACCACACGT TCATTCAAAT GAACTTCCGG	4260
	TTTAAATAT AATCCTTGCT TAATTGTATC TTTTAAATTT TCATGACAAG GTTTTGCGC	4320
	AATGCCGCCA ATATCCCAAA TAATTTTTTC AGGGTAAATT CTCATCTTAC CCCCTAATTC	4380
45	AGATTGAACA TCTATCAATC TTACAGACAT ATCTCGCAAT CCAGCATAAA AGCTTGATA	4440
	CAAACCAGAC GGACCGCCAC CAATGATTGT AACATCTTTC ATTATGTGCC TCCTATGACT	4500
50	CTCTATATTC ATTTCTTTCA TTAACGTGCT CAAATTGATA ATTATTATCA TTTAAAGCCA	4560
	TTATACTATT AATATTTATA TTGTTAAAT AAATCGCATA GTTAGCCATG AATTATCAAT	4620

55

GAAAGATGTG TATATTTTTT AGTTCTAGTT ATATTATTTT TTTAAAGACT CATCACGTGG 4740
 TTCTTTAAGA ATTGCTTGTC TTAAAAGGAA AAATAGCAAC AATAAACCTG CAAGCATACC 4800
 5 TGTGTGCCCCA ATACCTGCAA AGCCTGChAA TGCTTCTGGA GAGTATGATT TACCAGTGAC 4860
 TTGGAAGAAT CCTTTTGTC 4879

(2) INFORMATION FOR SEQ ID NO: 90:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1560 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

20 ATAATGTCTT AGaTTGATTG GGAGTTTTTT TAATTTTTTT GAAATTAAAT TAATCTGTAs 60
 yTAATAAAAA ATTTGAATAA CTGACACaYt TTTTGTATCA TAGCTAyATA CTTTGTGAAT 120
 TAATTCACAT TATAATAAGA GTGAAGATAA GAGTATTATA AATnATCTTT AAATAAATAT 180
 25 ATGTGAAGTA AAAATTACAC GTTAGCATAT CGATTATGgT CATTTCkTTT AACATATTAA 240
 CTgGGGaACG TTAAAAGTTA ACGGkTGATA TCyAACTAAA AACAAGGTCA CAGTAGTATG 300
 TTTTAATCTG GCGTCTATTA CAAATAAAAA TTACATCTAT AATTATTcGT TTTCTTTTTT 360
 30 GAAAGTAATA GCCAATTAAT ATCATAcata CTGGAGTGAC TATAAGGAGG ACATTATTAT 420
 GAGAGCAGCA GTTGTAACGA AAGATCACAa AGTAAGTATT GAGGACAAAA AGTTAAGAGC 480
 TTTAAACCT GGTGAAGCGT TGGTACAAAC GGAATATTGT GGCgTTTGTC ATACCGATTT 540
 35 ACATGTTAAG AATGCTGATT TTGGTGATGT TACAGGCGTT ACTTTAGGTC ATGAAGGTAT 600
 TGGTfAAAGTC ATCGAAGTTG CGGAAGATGT AGAATCATTa AAAATTGGAG ACCGTGTGTC 660
 40 TATCGCTTGG ATGTTcGAAA GCTGTGGAAG ATGTGAATAT TGTACAACAG GTCGTGAAAC 720
 ACTTTGCCGT AGTGTGAAAA ATGCTGGTTA TACAGTAGAT GGTGCAATGG CTGAACAAGT 780
 TATTGTTACT GCAGACTATG CTGTGAAAGT ACCTGAAAAA TTAGATCCAG CAGCAGCGTC 840
 45 TTCTATTACA TGCGCAGGTG TGACAACTTA TAAAGCTGTA AAAGTAAGTA ATGTAAaACC 900
 TGGACAATGG TTAGGTGTTT TTGGTATAGG TGGTTTAGGT AACCTAGCTT TACAATATGC 960
 TAAAAACGTT ATGGGGGCTA AAATTGTTGC CTTCGACATC AATGATGATA AATTAGCATT 1020
 50 CGCGAAAGAA TTAGGTGCTG ATGCTATTAT TAATTCTAAA GATGTTGATC CAGTTGCAGA 1080
 AGTTATGAAA TTAAGTATA ACAAAGGATT AGATGCAACA GTGGTAACTT CAGTTGCTAA 1140

TTTACCTGTT GATAAAATGA ACTTAGATAT CCCAAGATcA GTGCTTGATG GTATTGAAGT 1260
 AGTAGGTTCA CTTGTTGGTA CAAGACAAGA CTTACGTGAA GCGTTTGAAT TTGCTGCTGA 1320
 5 AAATAAAGTA ACACCTAAAG TTCAATTAAG AAAATTAGAA GAAATCAATG ATATTTTGA 1380
 AGAAATGGAA AATGGTACTA TAACTGGTAG AATGGTTATT AAATTTTAAA AATATCAACT 1440
 GACTATATAG ATAAAGAAGG TAGTGCTCTG AACACTATCA TTATTAATCA AACCCCGAGG 1500
 10 TTTTCCTGAA AAGATAGTGG nAAATCCCCG TGTTTTTTGG GTTTGAGGnG GTTGnTGTA 1560

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11014 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

GTCCTGTnGc TGCAATGAAT ACGCCTAAAA ATCCAGGGAT GTAATGGATA CTTTGTGGTA 60
 25 GTACTAATGA TAGAAATGAT AAAAATGAAA TCACAAAGGC TACGCTCGCA AAAGCTTGAC 120
 ATGTACGCTT ATCGCCATAA TCTAACCTG TACGTATATG TAATAAATAC TGTAATCCGA 180
 TACTTAAATA CATAATTGCC ACGCATAAGA AGAATGGGAA GAATGTCTTT TCAAAGTCCG 240
 30 GATATAGGCT GTTAGATAGG AAGACCATGA TGAACATATT AACATCATA AACGAGACGT 300
 CTTTGAATGT AACTTGACCA AATCGATTG TAAAAAATGT TTGATGAGAC CACATTAACC 360
 ATAAGAACAA ACTCATGACG ATGTATTTGA AAAATAAATC AGCTGAAATG GAACCGTTTT 420
 35 GTGTTGTTAA AATCACATGT GCAATTTTTT GAATGGCATA GACGAAAATT AAATCAAAGA 480
 ACAACTCATG GAATCCTGCA CGCTTTTCAG CTAAATGTTT TGGTGTTAAT GCATTAACCA 540
 TAAAATTTTA ACTCCTTTAA GATGTGTAAT TAATTTACTA AGTATACTAT TTATTTTTTC 600
 40 TAGTGAATAG GGGCAGATTT GGCATGAAG TGGAAGGAGA GGTGACTGCA AGGTAATTGC 660
 GGAATTAACA ATCATCAGCG ATTTAATATT TGA CTGGAGA CGTCATGGTA ATAAAAAATT 720
 45 GATGAGAAAT TGATGGTGAA ACCAGCTGTG AATAsCGaTG cAATGATrSA TAGaATTTAA 780
 TTAGAGTCAT TACGCGaAAT GATTAATGAT AATTGTGGT AAATCAAAGC aTAATTTTGT 840
 ACTATAGATG AGGATGATAG AGCATATTTA AGAGGGTGAA ATGTTAAAGT GAAACCGTTT 900
 50 ACGTTTCCGA TTGCCCAAAC AAATTACATC ATTGATAAT ATGATTTGTT AAATGCATAA 960
 CAAGAATGAA AATGTAACAT ACGTAGCAAT TGGTTTCATA AATTGGATGT TAGTGGCGTA 1020

	TGACGAGAGT	CGTATTAGCA	GCAGCATACA	GGACACCTAT	TGGCGTTTTT	GGAGGTGCGT	1140
	TTAAAGACGT	GCCAGCCTAT	GATTTAGGTG	CGACTTTAAT	AGAACATATT	ATTAAAGAGA	1200
5	CGGGTTTGAA	TCCAAGTGAG	ATTGATGAAG	TTATCATCGG	TAACGTACTA	CAAGCAGGAC	1260
	AAGGACAAAA	TCCAGCACGA	ATTGCTGCTA	TGAAAGGTGG	CTTGCCAGAm	ACAGTACCTG	1320
	CATTTACGGT	GaATAAAGTA	TGTGGTTCTG	GGTTAAAGTC	GATTCAATTA	GCATATCAAT	1380
10	CTATTGTGAC	TGGTGAAAAT	GACATCGTGC	TAGCTGGCGG	TATGGAGAAT	ATGTCTCAAT	1440
	CACCAATGCT	TGTCAACAAC	AGTCGCTTTG	GTTTTAAAAT	GGGACATCAA	TCAATGGTTG	1500
15	ATAGCATGGT	ATATGATGGT	TTAACAGATG	TATTTAATCA	ATATCATATG	GGTATTACTG	1560
	CTGAAAATTT	AGTAGAGCAA	TATGGTATTT	CAAGAGAAGA	ACAAGATACA	TTTGCTGTAA	1620
	ACTCACAACA	AAAAGCAGTA	CGTGCACAGC	AAAATGGTGA	ATTGATAGT	GAAATAGTTC	1680
20	CAGTATCGAT	TCCTCAACGT	AAAGGTGAAC	CAATCGTAGT	CACTAAGGAT	GAAGGTGTAC	1740
	GTGAAAATGT	ATCAGTCGAA	AAATTAAGTC	GATTAAGACC	AGCTTTCAAA	AAAGACGGTA	1800
	CAGTTACAGC	AGGTAATGCA	TCAGGAATCA	ATGATGGTGC	TGCGATGATG	TTAGTCATGT	1860
25	CAGAAGACAA	AGCTAAAGAA	TTAAATATCG	AACCATTGGC	AGTGCTTGAT	GGCTTTGGAA	1920
	GTCATGGTGT	AGATCCTTCT	ATTATGGGTA	TTGCACCAGT	TGGCGCTGTA	GAAAAGGCTT	1980
	TGAAACGTAG	TAAAAAAGAA	TTAAGCGATA	TTGATGTATT	TGAATTAAAT	GAAGCAITTTG	2040
30	CAGCACAATC	ATTAGCTGTT	GATCgTGAAT	TAAAATTACC	TCCTGAAAAG	GTGAATGTTA	2100
	AAGGTGGCGC	TATTGCATTA	GGACATCCTA	TTGGTGCATC	TGGTGCTAGA	GTATTAGTGA	2160
	CATTATTGCA	TCAACTGAAT	GATGAAGTTG	AACTGGTTT	AACATCATTG	TGTATTGGTG	2220
35	GCGGTChAAC	TATCGCTGCA	GTTGTATCAA	AGTATAAATA	ATAAGAAAAC	AGGTTATCAC	2280
	AACAStATTA	ATtACATGTT	GGCATAACCT	GTTTTTATTT	GTTTATGGAT	TTATTGGGTA	2340
40	ATATTAGTCA	TTTGATGGTT	TAATTGCAAA	TGCTCTAACA	GGGAACCCAG	GTGCATCTTT	2400
	TGGTTTAGGG	CTGATAGCGT	AAATGATGGC	GCCACGAGTT	GGTAATTGAT	CTAAATTAGT	2460
	TAATAACTCG	ACTTGGTATT	TATCCTGACC	AAGAATATAA	CGTTCGCCAA	CTAAATCACC	2520
45	ATTTTTTACA	ACGTCCACAG	ATGCATCGGT	ATCGAATGTT	TCATGACCAA	CAGCTTCAAC	2580
	ACGACGTCT	TCAATTAAGT	ACTTCAAAGC	ATCTAATCCC	CAACCCGGTG	CATGTTGTTG	2640
	TCCGTTTCGA	TCTTTGTTTT	CAAACCTTTC	AATATTAGGC	CAACGTTTTG	ACCAATCGGT	2700
50	ACGAAGTGCA	ACAAAAGTGC	CAGGTTCAAT	AGTACCATGC	TCTTTTTCCC	ATGCTTCTAT	2760
	ATGCGCACGT	GTTACGATGA	AATCATTGTT	GTTGCTACT	TCTGTTGAAA	AGTCTAATAC	2820

55

	AAAGTGAATT GGTGCATCAA TGTGAGTACC ATATTGCGTT ACAATATTCC AACGTTGCAC	2940
	ATAGAAACCA TGATCTTTAA CCGTGAATAA AGTTGAAACT TCGCCTTTTT CAAACTCACT	3000
5	AAAACGTGGT ATTTCCGGAT CAAATGTATG CGTTAAATCA ACCCAAGTTG CTTGTTTTAA	3060
	AGTATTTAAT TGTTGCCATA AAGGATATTG TGTACATAAA TCACCCGTTT TTAGTTTATT	3120
10	ATATGATAAA TGCTGCGATT ATTCTTGGCG TTTAGCTTTA ACAGCATTCA CAAGCACAGT	3180
	CAATGCATCT TTAACCTCTT CTTCTTTTCG CGTTTTTAAA CCACAGTCAG GGTTTACCCA	3240
	GAATAATGAG CCGTGCATTG GTTGTAGTGA ACGATTGATT GCTGTAGTAA TTTCTTCTTT	3300
15	TGTTGGAATA CGTGGACTAT GAATATCATA TACACCTAGA CCAATACCTA AATCATAATT	3360
	AATATCTTCA AAGTCTTTAA TTAAATCACC ATGGCTACGA GATGTTTCAA TTGAAATAAC	3420
	ATCAGCATCT AAGTCATGAA TAGCATGAAT GATTTGACCG AATTGAGAAT AACACATATG	3480
20	TGTATGGATT TGAGTTTCAT CACGAACTGA AGACGTTGCA AGTTTAAATG ATAAAACAGC	3540
	ATCTTTAAGA TATTGTTTCG GATATTCAGA GCGTAATGGT AAGCCTTCAC GTAATGCAGG	3600
	TTTCGCAACT TGGATAACTT TGATTCTGCG AGCTTCAAGT GCTAATACTT CTTGCTTGAT	3660
25	TGCTAAAGCA ATTTGATCTT GAACGACTTT ACGTGGTAAA TCAACACGTT CAAATGACCA	3720
	GTTTAGAATT GTTACAGGTC CAGTTAACAT ACCTTTAACT GGTTTATCTG TTAAGCTTTG	3780
	TGCATAAACT GTTTCATCAA CAGTTAAAGG CGCTGTCCAT TTTACATCAC CATAAATGAT	3840
30	TGGTGGTTTT ACGGCACGTG AACCATATGA TTGCACCCAA CCGAATTTAG TTACTAAGAA	3900
	ACCTTGTAAT TTTTCTCCGA AGAATTCAAC CATGTCATTA CGTTCAAATT CACCGTGAAC	3960
35	TAATACATCT AAGCCAATGT CTTCTTGAAT TTTAATCCAT CGAGCAATTT CATTTTTTAA	4020
	GAATGTTTCA TATGCTTCGT CTGTAATGCG TTTGTTCTTC CAATCTGCAC GGTATTTTCG	4080
	AACTTCTCGG CTCTGTGGGA ATGATCCAAT AGTTGTTGTT GGTAAATCCG GTAAGTTCAA	4140
40	ACGTTTTTGT TGTTGTTCAA TACGTTGCGC GAATGGTGAT TGTCTTGAAG TACGCACGCT	4200
	TTCGAAATCA TAATCTAAGT TTTTGAATGA TTGATTTTGG AAACGCTCAT AACGTGCTTT	4260
	TAATTTATCA TATTTAACAC TATCGTTTGG ATTAAATAGG CGACGCAATG CATCTAATTC	4320
45	GTCTAATTTT TCAGTTGCAA AGCTTAAGCC TTCGCCAACA CTGTATCTA ATGTTTCATC	4380
	ATCTAAAGAT ACTGGAACAT GTAATAATGA AGATGATGGT TGAATGACAA GTTCATTAGT	4440
	GTGTGCTAAC AATTTATCGA TTAAGACTTT TTTAGCTTCA ATGTCACTTG CCCATACATT	4500
50	ACGACCATCA ATAATTCCAG CGTATAATGT TTTTGATTTA TCAAAATCTC CAGCTTCAAT	4560
	TTGTTTAAGG TTATAGCCAT TATCATGGAC AAAGTCTAAA CCTATACCAC CAACAGGTAA	4620

55

	AACACCAGCT	TTTTCGAAAT	AGTCATAAGC	TTCACGTGTA	ATATTTTCAT	AGCTTTTCGCT	4740
	GTCGTCTGTA	ACTAAGATTG	GCTCATCAAC	TTGAATGTAC	TCAGCACCTG	CATCAATTAA	4800
5	TGATTCAAAC	ACTTCTTTAT	AAAGTGGTAA	TAACGTTTTA	ACTTTTCTT	CAAAAGTTTG	4860
	GTGACCGCCT	TTTGATAATT	TAACAAAAGT	AATCGGACCA	ACAATGACAG	GGTGAGCGTT	4920
	AACGTTTAAA	GATTGGGCAT	ATTAAAGCG	ATCTAATAAT	ACATTGCGAC	TCACTTTAGG	4980
10	CTCAACATTG	TCCCATTGAG	GTACGATGTA	ATGATAGTTA	GTGTTAAACC	ATTTTATAAG	5040
	TGCACCTGCA	ACATGGTCTT	TATTACCGCG	AGCAATATCA	AATAATAAAT	CATCATCAAT	5100
	AGTTCTTCCT	TGGAAACGTT	CAGGGATGAT	GTTGAATAAT	AATGACGTAT	CTAATATATG	5160
15	GTCATATAAA	GAGAAATCAC	CAACTGGGAT	GCTATCTAAG	TGATAGTACT	TTTGtAATAA	5220
	TAAATTlyCT	TTATGTAGAT	CAGTTAATGT	TTGATCTAAT	TCTTCTTTAG	AAATCTTCTT	5280
20	TGCCCCAATA	CTTTCGATGG	CTTTTTTCCA	TTCTCTTTTT	CTACCTAATC	TTGGGAATCC	5340
	TAAGTTTGAT	GTTTTAATTG	TTGTCATAAT	ATTGCCTCCT	TGTGAGCAGT	AATAGATTTT	5400
	GAGTATGCTG	CAAGTTCTAA	TGAATCTTCG	ACATTTTGAA	ACGGTGTGAT	AATGTATAAA	5460
25	CCATTAAAAAT	ATTGATGAAC	AGTATCGATT	AAATCCTTTG	AAAGCTTAAG	ACTTAGTTCT	5520
	CGTGTTTTGG	CTTTATCATC	TTAACTGCT	TCAAATTGTT	GTAAAATTTC	ATCTGACATC	5580
	TTGATTCCTG	GCACCTCATT	ATGCAAAAAG	AGTGCGTTTT	TGTAACCTGC	GATAGGCATA	5640
30	ATGCCTATGA	AAAATGGTTT	GTTCAAGTGC	TTAGTGGCAT	GGTAAATTTC	AATGATTTTC	5700
	TCTTTGCTGT	ACACGGGTTG	TGTTATAAAA	TAAGACATTC	CGCTTTCTAT	CTTTTTCTCT	5760
	AATCTTTTGA	CGGCACCATA	TAATTTACGA	ACATTAGGGT	TAAAGGCGCC	AgcGATGTTG	5820
35	AAGTGTGTAC	GTTTCTTCAG	CGCATCACCG	TCAGTGTTAA	TACCTTGATT	AAATCTTAGA	5880
	GCGAGTTCAG	TTAATCCTTT	AGAATTAACA	TCATAGACAT	TGGTTGCACC	TGGTAAGTGA	5940
	CCAACTTTTG	AAGGATCACC	AGTTATGGCT	AATATTTTCG	TAACGCCAAT	GAGCGATAAT	6000
40	CCAAGTAAAT	GGGACTGCAA	GCCGATTAAG	TTTCGGTCTC	GACATGTAAT	ATGTACGAGT	6060
	GGTTCAATAT	TGTAATATTG	CTTAATTAAG	CTAGCAGCAG	CAATATTGCT	AATTCTGACA	6120
45	GTTGCCAATG	AATTATCTGC	GAGTGTTACC	GCATCTACAT	TAGCTTTATC	AAGTTTAGCG	6180
	ATATTTTCAA	AAAATCTATC	CGTGTCTAAA	TGTTTCGGTG	TATCCAATTC	GATAATAACG	6240
	GTTGGACGTT	CTTGAACCTT	AGATGTTAAT	GATTGTCTAA	CTTTATTTTG	AGATGGATTG	6300
50	AAAAGTGCTT	TCGTTGGTAT	CGGAATCACT	TTTTTGTCAT	TAACAGGTTT	AAGTGTCTGA	6360
	ATAGATTCTT	TAATAAATTT	GATGTGCTCT	GGCGTTGTAC	CACAGCAACC	ACCAATTAAA	6420

55

	TACTTAAATT CACTATTTTC AATATCTAAT AAGCTGGCAT TTGGATAACA AGATAAGAAT	6540
	GCGTGCTCTG GTAATTCAAT ATGTGTGAAA GACTCTTGCA TATGGTGCGG GCCATGATGA	6600
5	CAATTGAGTC CCACGATGTT TGCACCACAT TGAACGAGTT GTTTTAATCC TTCATTGATT	6660
	GCCTGACCAT TAACTAAGTA ATTTGTGTTT GAAGCGGTTA ATTGAGCAAT GATTGGAATG	6720
	TCGTATTTCT TTCTCGTTTCG TGAAATGACA TTTGTTAACT CTTCTAGGTC GTAATACGTT	6780
10	TCGAAAAGTA GCGCGTCAAC GCCTTCTTCA ATTAAGGTGT CTATTGAAT TTCAGTATGA	6840
	TAAAGAATAG TTTGTAAGCT GATATCCTCT TGTTTGATAC CTCTAAACCC ACCAACTGTG	6900
	CCTAATATAT ACGTATCTTT ATTTGCTGCT TTTTTGCGA TCGGAACGGC GGCTTGATGT	6960
15	ATTGCTTTAA CTTTATCTTC AAGACCGAAT CGTTTTAACT TTTCAAAAT TGCACCATAA	7020
	GTATTGGTTT GAATGACATC AGCACC GGCT TCAATATATG AACGATGGAT GCGTTCAACT	7080
20	TTATCTGGAT GGCTAAGATT ATATGCTTCT GGACAGGTGT CTAATCCTTC AGAGTATAAA	7140
	ATGGTTCCTA TAGCGCCATC AGCTACTAAA ACATTATCTT TCAATGTGT GAGGAATTGA	7200
	CTCATTGAAT GCCTCCTTTA ATGCGTATTT GATGCTGCA ATGAGTTCAT CAGGATCTTC	7260
25	GAGACCAACA CTTAATCGGA ATAGACCGAA AGTGATACCA CGTTCCTGTC TCACTTCTTC	7320
	AGGTAGTGCA GCGTGAGACA TTGTTGCTGG ATGTGAAAGG ATCGTTTCAA CACCGCCCAG	7380
	ACTCACTGAA ACGAGTGGTA ATGTCAGTGC ATCGACAAAT TGTGTGCTT TAGACTCATC	7440
30	AGCTAAACGA AAGCCAATAA CGGCACCGCC ATTTTGTAGCT TGTCTAAAT GAGCAGTAGT	7500
	GAGTCCCGGA TAATAAATCT CTGAAATTTT ATCTTGCTTT ATTAAAAATG ACACGATTTT	7560
	TTGAGCGTTT TCGACAGATT GTTTAAATCT GATTGGAAAA GTTTTAAAT GTTTAGCAAG	7620
35	TGTCCAGCTA TCCTGAGCAG ATAACATATT GCCTGTACCA TTTTGTATTA AATAAAGAGC	7680
	GTCCTAATT GCCTCATTAT TAGTTATGAC AGCACCAGCA ATTAAATCGC TATGTCCACT	7740
40	TAAAAATTTT GTAGCACTAT GAATGACAAT ATCAGCGCCA AGTAATAAAG GTGATTGACC	7800
	TAACGGTGTC ATAAATGTAT TGTCCACAGC TACCAGTAGT TCATGCTTTT CGGCTATTTT	7860
	AGAAACAGCT TTGATATCAG TAATTTTAAA ACAGGGATTC GATGGTGTTT CGATATAAAT	7920
45	TAATTTTGTG TTTGATTGAA TGGCACCCTC GATTGTTCG AGCTTTGTAG TATCTACGGT	7980
	TGTAAATTCA ATATTAAATC GATTCAAAAT TTGCTCAGTG AGGCGAAAAG TACCGCCATA	8040
	TACATCATCG GGTAAGATGA CATGATCACC AGATTTGAAA GTCAAAAGTA CTGCTGAAAT	8100
50	AGCAGCAATA CCTGATGCAA AAGCAAAAGC GAATTTTCCC TGTCTAATC GTGCTAACTT	8160
	CTCTTCTAAA AGTTCACGGT TAGGGTTGCC CTTCTGTCAT AATCATATTT AACATCGCCA	8220

55

	TCCACACCTC	TACGCCAATC	GAATATCACT	TCTGTCTCTT	TTGAAAGTGT	CATACAATCT	8340
	CTCCAATCTG	AGCTTTATCT	AATGCTTGGA	TGATATCGCG	TTCGATGTCT	TCATAATTTT	8400
5	CAACACCTAG	TGATAAGCGG	ATTAAATACT	CATCAATGCC	ACGTTTATCT	TTTTCAGCAT	8460
	CTGGCATATC	AACATGTGTT	TGGGTGTAAG	GGAAGGTCAC	TAATGTTTCA	GTACCTCCTA	8520
10	AACTTTCTGC	AAAAATGCAA	ATGTCTAAAT	TTTCTAATAA	TTTAGCGACG	CTATAGGCCT	8580
	TGTTAAGTCT	TAAACTAAGC	ATGCCAGTTT	GCCCCGTATA	TAGTACTTCG	TCAATTGCTT	8640
	GAAGTGACTG	ACATTTTTTA	GCAAGTTTTT	TAGCGTTTGA	TTGCGCACGC	TCAATGCGTA	8700
15	AATGCAAAGT	TTTAAGTCCA	CGTAACAACA	AATAACTATC	TATTGGTGAA	AGTGTGCGC	8760
	CAGTCATGTT	GTGAAAATCA	AACAACGTGT	GCGCGAGTGA	TTCATCTTTG	ACGGTTACGA	8820
	CACCTGCTAG	TACATCGTTA	TGTCCGCCAA	TATATTTTCG	GGCTGAATGT	AAGACTATAT	8880
20	CAGCACCTTC	TGCTAGTGGT	GTTGAAAGAT	AAGGTGTTAA	AAAAGTATTG	TCGATAATTG	8940
	ACAATAAGCC	TTTAGCTTTA	CAAAGTTGAT	AGTATGGCTT	TACATCAATA	GCAATCATTT	9000
	GTGGGTTAGA	TATTGGTTCA	ATGAATAATG	CAACTGTTTT	ATCAGTGATT	TCTTTTTCAA	9060
25	CTTGTTTATA	ATCTGTAAAA	TCAACGTACT	TAAATTTGAT	ATCGTATTGT	TGCTCGTAAA	9120
	ATTCAAATAA	TCTAAATGTG	CCACCATATA	AATCGAATGA	AACTAAAATT	TCATCATGAG	9180
	GTTTAAATAG	ATTACATATT	AATTGAATGG	CTGACATTCC	ACTTGATGTA	GCGAATGATG	9240
30	CAATACCATG	CTCAAGTTTG	GCAAAACAGG	TTTCAAATGT	TGAGCGTGTA	GGATTTTTAG	9300
	TACGTGTATA	ATCAAAACCT	GTCGATTGTC	CTAGTTTTGG	ATGCTTGTAG	GCAGTAGATA	9360
35	AATGGATTGG	ATTCGCTATA	GCACCGGTTG	AATCATCGGT	TAATGTGATT	TGGGCTAACT	9420
	GTGTATCCTT	CATATTAAGA	CCCTCCTATA	AGAAAAATA	AAAAAAGCTT	CCGTCCTTCG	9480
	TACCCGAATG	AATCGGATAA	AAAGGACGAA	AGCTTATGTT	TCGCGGTACC	ACCTTTATTT	9540
40	GTTATTCCAT	CGCTGAAATA	ACCTTATTCA	GTACGCATTA	AAAGTAAATA	TGCTTACTGA	9600
	ACAATTATCA	CAATTAAAGT	CAGTAAGTAA	GGATATAGTA	ATGTGCTATC	CCATACTTAT	9660
	TAACAAAAAA	TCGTGCGTAA	AGAATCCAGT	ACGCCATTTA	ACATCAATGT	TAATACTGTA	9720
45	TCGCTATAAC	GGGCGAACCC	GTAGACACCT	CATATTGGCA	TCAACACTCC	AAGGCCATTT	9780
	TCAAACACGC	TTTCAAAATC	TTCTCTCAGC	TACTAAAGAC	TCTCTGTATA	AGCAGGGTGT	9840
	GTTTTACTTy	CCTCTTTATT	GTGTTTACGT	TTCATTAAAC	TGTTATAAGA	TATTAATTAG	9900
50	CTTACAGAGT	AAAAAAGAT	TTGTCAACAA	TTATTCAGAA	AATTTTGATT	TAAAAGTTAA	9960
	TTTGTTTGTG	AAATTGTAAT	TGGTATCTTG	AAGTTGAAAA	ATGAATTATT	TTTTAAATAA	10020
55							

TCAAATAAAA AGTGATGTGA GTGAATTGTC AAAAAGTGAA GATCAACGTA TTAATAAAAC 10140
 AAAAGATGAA CAAATTAAGC AAATAGATAT ATCGGATATC AAACCGAATC CGTATCAGCC 10200
 5 CCGAAAAACT TTCGATGAAA ATCATTTAAA TGATTGCGCA GATTCAATTA AGCAATATGG 10260
 AATTTTGCAA CCAATTGTGC TTAGAAAAAC AGTTCAAGGT TATTACATTG TAGTTGGTGA 10320
 AAGAAGGTTT AGAGCTTCGA AAATGCTGG TCTAAAATAC GTATCAGCGA TTATCAAAGA 10380
 10 TTTAACAGAT GAAGATATGA TGGAAGTGGC GGTTCATCGAA AATTTACAAC GAGAAGACTT 10440
 AAATGCGATT GAAGAAGCTG AAAGTTATCA ACGTTTGATG ACAGATTGGA AAATTACACA 10500
 ACAAGAAGTA GCGAAACGAT TGAGTAAGTC GCGCCCGTAT ATAGCGAATA TGTGAGGTT 10560
 15 ATTACATTG CCGAAAAAGA TTGCTGACAT GGTAAAAGAT GGGCGACTGA CAAGTGCACA 10620
 TGGACGAACG TTATTGGCAA TTAAAGATGA ACAACAAATG CTTAGGTTAG CGAAACGGGT 10680
 TGTTAAAGAA AAGTGGAGTG TCAGATATTT AGAAAACCAT GTTAATGAAT TAAAAAATGT 10740
 20 TTCGTCAAAG TCGGAAACAG ACAAGTAGA TATAACTAAG CCTAAATTTA TAAAGCAGCA 10800
 AGAACGACAG TTGCGAGAAC AGTATGGTAC CAAAGTAGAT ATATCAATAA AAAAATCGGT 10860
 25 TGGTAAAATC TCATTTGAGT TTGATTCACA AGAAGATTTT GTGAGAATAA TTGAACAATT 10920
 AAATCGTAGG TATGGTAAAT AGTTACACAA TTTTATATAA TAACTCTTTG TGCAAGTGTA 10980
 AATAAATTGT AATCAGTGAC ATTTGATTCT AGAT 11014

30 (2) INFORMATION FOR SEQ ID NO: 92:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6022 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

40 TCCCCTTATG GAATTCACA TTCTAGTTTA CATAATATAT ATTATAGGAA GTTATATGTG 60
 TGTAACGCAA AAGGTACCCT ACATCATAAT CATTATCTAA TATCGTCACA TAACTTACTT 120
 45 ATGCTATAAT CATGGTATTA TATTGTTTGG AGTGATTGGA TGAGATTGT CTTTGATATT 180
 GATGGTACGC TTTGTTTTGA CGGCCGATTA ATTGACCAGA CTATTATTGA TACATTGTTA 240
 CAATTACAAC ATGATGGTCA TGAAGTTATA TTTGCATCAG CACGTCCGAT TCGTGATTG 300
 50 TTGCCAGTTT TACCATCAGT ATTTTCATCAG CACACATTAA TTGGCGCAA TGGTGCTATG 360
 ATTTACAGC AATCAAAGAT TTCTGTTATC AAACCAATTC ATACTGATAC ATATCATCAT 420

55

	GCTGCACAAC TTGACGCTGn AGAACGCGAT TTTTGAGCGT TTAGATCCAC ATAAGCTGGC	540
	CAGTTGTATT GATGTTGCAA ATATCGACAC GCCAATCAAG AkTATTTTAT TAAATATAGA	600
5	CCCCGCACAA ATTACAAC TAATTAGACGA GCTAGATAAA TACCATCAAG AATTGGAAAT	660
	GATTCAACCAT TCAAATGAGT ATAACATTGA TATAACAGCG CAAAATATTA ACAAATATAC	720
	TGCATTACAA TATATATTTG ATGCAGATGT TAAATATATA GCATTTGGTA ATGACCACAA	780
10	TGATATTGTC ATGTTACAAC ATGCTAGTAG TGGCTATATT ATAGGACCAT CAGAAGCATA	840
	CACACACGCA ATATTGAAAC TTGATAAAAT CAAACACATC AATAATAATG CACAAGCTAT	900
	TTGCAAAGTC TTAAATCAT ATAAATAAAA ACACCCCTAT CAAATGATAA TCATTATCAA	960
15	TCGATAGGGG CTATTTTAAT AAAATTCGTC CTCGAACATT TCTTCCTCTT CATCTAATCC	1020
	AAATAATTCT GCCATTTCTC CATGTTCAAT TAACATGTTT AAATATGCAT CGCGGAGTTC	1080
20	TTCTTCACTC ATATCATTAA TCATTTCTTT AAGACTATCA ATCCACATAT TTCTGCGTAA	1140
	TTGATAGTCT TCTTCAACTT CGTTTAACAT CATTATATGT TTATTTGCTG CTTCTGGACT	1200
	AGCTGTAAAG AGTAATGCAA TCATATGTTT ACATATCACT CGTCTTCCAT CAGCATGAGG	1260
25	ACAATTACAT ATGGATTTTC TAGGATGTTT CATATCAATA TAACAACGAT ATACTTTGTT	1320
	GCCACTGCCC TTTACTTCAG CCTCATGCTG CGTTTCTGAA AATGATTTTA AGTTAATGAC	1380
	GCATTCACCT TGATAATAAT TAAAGCCTCT TTCTATAGAA CGAATACTTG CAATATCAAG	1440
30	TAATCCCAT AATGaTACTC CTTTTATTA TTATTTTAA ATAAAGaAA TAAATAGAT	1500
	AAGTGTCTAG ATTAAAATAC TTGATTTATC TATATTTTAT AACAGTCTA GAATTATCGC	1560
	ATTCTTAAAT AACTAATATG AAAATGcTTG CACTAATTct TTTGTATAAG GGTGTCTATC	1620
35	AACATTAAAT AATTCctCTA TTGCAAAATC ATCGACTATC ATGCCATCCT TAAGAACGAT	1680
	AATTCTATTA ACTAAGCGTT GTAACACGGA TAAATCATGA GAAATAACGA TAAATGATT	1740
	TAAGTTCGTA ATCGTTTGCG CTTTTAATAT ATTGATTACA TTTTGTTTCAG CTATAACATC	1800
40	TAAATTTGAA GTTATCTCAT CACATATTAA AACGCGAGGC TGTGCTAATA ACGAACGCAT	1860
	GACATTAAAT CTTTGTAATT GTCGCGCACT CACTTCGCTT GGTAATTTAG TCAATAATTG	1920
	CGCGTTTAAAC TCAAAAGTAG ATAAATGTTG TAATAATAAT TGATCCTGAG CAGTATTATC	1980
45	AGTTAGACCT CTGTAATAAT ATAACGCTTC TTTTAATGAG GTCTCAATCG TCCAATCAGG	2040
	GTTAAAGCTA GTTAAAGGGT GTTGGAATAT CCGTAACACA GCATTGTCAC TTAAGTAAAT	2100
50	CTCTCCTTTA ACAGGTTTAA ACAAGCCAAG AACCAATGAA GCGAGCGTAC TTTTACCACA	2160
	GCCACTTTCG CCTAAAATAC CAACATTTTC TCCATCAGGT ATAGTAATAT TGATATCTTG	2220

	CCCTCTTTAA TTGTGTTCTA TATTTAATTA GACGTTTCAGT ATACGGATGC AAATGCTCAT	2340
	ACTTGAAATG ATTAATATTA CCTCGTTCAA TGATTTGACC TTCTTTTAAA ACATAAATGT	2400
5	ACTGACAATA TTTCAATACA TGACTTAAGT TATGTGTGAT AATAAATAAT GTTTGACCAT	2460
	GTTCTAATAC AATATGCTGT AATAAATCCA TCACCTTGATT ACCGTTTCAA GCATCCAATG	2520
	ATGCAACTGG TTCGTCTGCA ATGATTAATT TAGGCTCCAA CATGAGAACG CTTGCTATGT	2580
10	ATACGCGTTC AAGTTGGCCC CCAGAAAGTT GGAACTATA TTTATTTAAT ATATCTTTGC	2640
	TTTGTAATTT AAGCCACGAC AAAGCCTTAT CAACCTTGGA CAAAGCCTCT TCTTTACTAC	2700
	CTTTATAATG CTTACGATAA ATCGCAGTTA ACTGTTTACC TAATTTAGTA TGGTCGTTAA	2760
15	AACCTTCTGC ATAATTTTGA GAAATATAGC CAATGTATG ACCATAATAT TGACTCAATC	2820
	TACTAACATT TTCCCATCA AATTGGTACG AATCATACGT GCAGCTTAAA TCAAATGGTA	2880
	AATATTCAAG TAAAGCTTTA GCAATCAAAC TTTTCCAGC GCCGCTCTCT CCAATCAAGG	2940
20	CATTAATCTG TTGACTAAAA ATTTTCAAAT CAATCCCTTT AATAAGAGAT TTCTCACTAG	3000
	TATCTTTTAT TGTAAATTT TGTATATCAA TGAGACTCAT CATATTCACC CCGTTGTTTC	3060
	AGCAATCTAT CTCTTAGTGC ATCACCAGTT AAATTAAAAA TTAAATAGT TATAGCAATG	3120
25	ACTGAAGCAG GTGCAATCAA CATAATTGGA TGAGACGAAA TAAATCACC ACCTTGTGTC	3180
	AACATAGCGC CCCaCTCTGG TGTGGCGGT TGTGCACCTA ACCCAATAAA TGATAGTGAA	3240
30	CTTATATATA GAATGATTTT ACCGAAATCA ACGACCATCA AAACGATAAT AGCCGGTATA	3300
	ATTTTAGGTG TTAAATGACG TATTAATATT GTTCTTGTTG GTACATGAAA TAATTGTGCC	3360
	ATTTTATAT AAGGCTTATT CATTTCGCTA TTAATATAC TTCTAGTCAA CCTTGTGTAA	3420
35	TTCATCCATT TTATTAATGT AATTGAGATA ACTAAATTCC ATAAAGATGG TTGAAAAAA	3480
	CTTGCTAAAG CAATCATGAT GATAAATTCT GGAATACTTA GACCAACATC AATAAACCTT	3540
	AACACTAATC GTTCAATCCA CCCTTTTTTG TATCCGGCAA ATAGACCTAG TGTAACACCT	3600
40	ATGACAACGA TAGCTATTAA TGTAAAAACA GTAACAAACA ATGTTGAACG TGCACCGATA	3660
	ATAATTCGGG TAAATAAATC TCTCCATAA TCATCAGTTC CTAATAAATG CAACCAACTA	3720
	ATAGGTTCAA AAGTTTGTGA TAAATTGACT TTGGTTGCAT TTTCACTACT GACAAAGAAT	3780
45	TGCAGTACAA TTACCACAAA AATAAATGCA ACGAATACAA AAAATATCAG GTTATTCTTT	3840
	GAAAATATTT TATGCATGAC GGTCACTACT TTCTGATATC AATGGTGTAT TGGTTTTGAT	3900
	TTTTGGATTT CCTAATTGTA AACGCTGCTT CGGATCAAGT AATAACGTTA ATAAATCAGC	3960
50	AATCGTATTG ATAATAACAA CGAAGAAGCC AATAAATAAC ACGCATCCTT GAATAACAGG	4020

55

	ATTTTCAATC ACTACAGTAC CACCTATTAG ACTGCCAAGT GAAATCCCTA GTAATGGGAT	4140
	AATCGGCAAA ATTGTTGGTT TTAGTAAATC ATGAATTAAA ATATAACGTT CATTTCATACC	4200
5	GCGTAATCTT GATGCTTGTA CGATATTACT TTGCAATAAC ATCAATAAAT TAGAACGCAC	4260
	TAAACGAATG ATGTATGCAC ACATACCTAA AGATAGCGTG ATTACAGGTA ATATAAACTG	4320
	ACTTAGTATA ACGCTATCTA TATTCATTAA ATTTGTGACA ATAAATAATA AAATAATACC	4380
10	GATAAAGAAC GCTGGTAAAC TAATCGATAG TGTGAGATC ACTCTAATCA CTTTATCCGT	4440
	CCACTTATGA AATCGTTTGG CTGCTATAAT GCCGAGCGGT ATAGATATGC ATAACGACAC	4500
	TACTAATGTT GAAAATGATA TGAGTAATGT TATGGGTGCA TAGTTGAATA ATATCTGTGT	4560
15	TACCGGTTCT TTTGATTCAA AACTTTTTCC TAAATTAAAA TGTAATAAAT GATTCATCCA	4620
	ATGCCACCAC TGTACCAATA AAGAATCATT TAATCCCAAT TTATCTTTGG TTGCATTTAT	4680
	TTGTTCCGTC GACACTTGTG CTACATCAAG ATGTAATATT TTATCAACAG GATTGCCTGG	4740
20	TGATAATTTT ATTAATGA ATGTAAGTGT AGAAATAACA AATAAAACAA CTATCATTTG	4800
	CATCAGTCTA TACAACATAG ACTTTATTAT GAACATAATA GTCCCCCTCC TTGTGTAAGT	4860
	TACTAACACT TTCTTTTAC ATGAGAATGG CGCATGTATA TGCAACTTAC ATATTAAGAA	4920
25	CTAACGTTCA TTATAGTATT ATCCATAAAG AAATTGAAGT ATATTTAATT TTTTAACAAA	4980
	ATCATTATAA AATATAATAT TTTGAATCAA GTCAACCATG TAAAATATAA AAAAGTCAAA	5040
30	ACAAAAACAA CTATAGCACT GTATTCCATC TCTTTCGAAA TAATTGTTAC TGCAGTGTA	5100
	CTTAAAAGTC GATGATTTTG TGCATATAGT TGTCGAATAT TATTTTTTAT CTTTACGGCG	5160
	AAGTTCAGCG CCTCATAGC CGTATTTTTC AATTGCTTT TCTAATTAC GCGCTTTTCT	5220
35	TTCTTTACGC CAATTTCTAG TAAATACCA TAATAGAAA CTAATTAATA AACTCATAAT	5280
	CGCTAAAAAT GCAGCGTATC CTAATAATGG TTGATATTTT ATATCTTGAA AATTGGAAT	5340
	AAAAAATGCA AGCACACCTA ATATAACAAA TGTAATTACT GCAGATACAA ACCATTTATT	5400
40	TAAAACTAAG CAACAGAATA TTGTTAATAA AATCATTATT AATGTTGTGA TCCATAAATA	5460
	ATTAGGCATA TCGAATAATG TCATATTCAT TCTCCTTTTA TTTCAATTACT TTCCTTGTAT	5520
	ACATTTTATT ATAAATTTT AAAAECTTAA ACAATAGCAG TCAGTTTCAA GCAATATTCT	5580
45	ATCTACTAAT AGAAAAATCA TTGTTCCCTG CGACATGGAA ATCGTAACAT TATCGTTTAG	5640
	GAGACAAAAT TATGTATAAT GAATGTATTA TACCAAAGGA GTGATTATAT GTCTCAAGGT	5700
	TTACCTTTAA GAGAAGATGT TCCTGTTTCA GAAACATGGG ATTTAGTAGA CTTATTTAAA	5760
50	GATGATCAAC AATATTATGA AAGTATTGAC GCTCTAGTAC AACAAGCAAA TCAATTTTAT	5820

55

GAAAATATTT TAATTGCCTT AGATCGCTTA AGTAATTATG CAGAACTACG TTTAAGTGTA 5940
 GATACTAGTA ATATCGAGGC ACAAGTATTG AGCGCTAAAT TATCTACTAC ATACGGTAAA 6000
 5 ATTGTTAAGC CAATTATCCT TT 6022

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 476 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

CCATCAATAA TGTATACATG ATTGGCATCA TATCCCCCTT TAATTAGAGA GCTACGTACA 60
 GTTTGTyTTA TTAAAGTAGA ACTAATAAAT AACCATCTCT TATGTGCACA AACACTTCCC 120
 GCAACAATTG ATTCAGTTTT ACCAACCCGT GGCATACCTC TAATGCCAAT CAACTTATGA 180
 CCTTCTTCTT TGAACAATTC AGCTAAAAAG TCTACTAACA AGCCTAAATC TTCACGCTCA 240
 AATCGAAAGG TTTTCTTATC TTTTGCATCT TGCTCAATAT ATCTTCCATG TCTTACTGCA 300
 AGACGGTCTC TTAATTCTGG TTTTTTAAGC TTTGTTATTT CAATTTTCATT TATACCACGA 360
 GCTATTTGCT CAAAACGTTT AACTTTTTC AATTGTCTG TTTTAATTAA AAGGCCTCGT 420
 30 TTACCTTGAT CAACACCAAT AATTGTAACA ATACTTATAC CTAACATACC TAATAA 476

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3633 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

AGAAATACAA CGAAGCATAT AAATATAACC GATCTTTTTT CTAATTGAAT ATTAAGTAAG 60
 TGTATGTA CTCTGGAAGT AGCACCTAGT rGGATTGTtC CTCCTACAAC AGGCCAAAAA 120
 45 TTTTATTTTT TAACTGGCTT AACAGTGTTC AGTTTTTCAT ACTCTTCTCT ACTAATTTTG 180
 GCGCACCTTT TTGGAATGAA CCAATTAATA AATGGAAAAA AGTATACAAG CCAAGTTCTT 240
 ATTACATCGA CCATTAAATA CTCATCATCA TACTTAATAA CTCTGTATTT CGGATTTTIA 300
 50 TTGATAATTT CGGTTTCACA AAGCAATAAT TATCACTTCC TATTAATAAC AAATTCACAC 360

	TTATATGACC	TTAAATATAT	AACATGAATC	TTTTTGTCTA	TTATTGAAGA	CATATTTATA	480
	AAGAAAAATA	GCATTGTCAT	AATAACCCAA	GCAATAAATA	CTATAATATT	TTGGATAGAT	540
5	AAACTAATCA	TTACATCTAA	GAACATGATT	GATAATCCAC	CACAGAAAAA	ATAAGAAAAT	600
	AGTACAAAGC	AAAGATTCTT	GAATGATGGA	AAAATCATAA	TTTTTCCATT	GCTACTCCGA	660
10	TCATTATAGA	TAGATAACTT	TACTTTCTGA	TTTAAATATA	TATAAAACAC	TAGAATACTT	720
	AATAATAAAA	CCGAACAAAT	GATAATAACG	CAATTTTTTT	CTAAATGAGA	ATCAGGTATA	780
	TATATTTTAT	CTCTAAACAT	AGTGCCAAAT	AAAAGTATGC	TACCTATAGC	TGGCCATAAA	840
15	GCTTTaTTTT	TAACTGGTTT	GACAATATTT	AAATTATCAA	AATCTTCTCT	GCTGATTTGG	900
	ACATATTTTT	TTGGTATTAA	CCAATTAATA	AACGGAAAGA	ACAAAATAA	CCAGGTGCTT	960
	ACTAAATCAA	TCATCAGATA	GTCGTTTTTA	TATTTAATAA	TTCTATATCT	GGGATTTTTG	1020
20	TTTACAACCTC	TAACCTCGCA	AAGCAATATC	TCCACTTCCG	TCTCGTTGGT	TTTATATCTA	1080
	ATACACTTTC	AGATACTTTA	TAAGTGTTTT	GTAATTTAGT	AACATACTAT	TTTCCTGTTT	1140
25	ATTACTTAAC	TTACGAACTA	CAATCTAAGT	TTAGTAATTT	CTATTGCTTT	TTAAGTTTGG	1200
	CATAAACCTT	TTTATTACTA	ATTGAGCCCA	TGCTTATTAG	AAAGAAAAAA	ATTGTAATAA	1260
	TAATCCACAT	AATAAATACC	AGTAGATTTT	GAGGTTTTAT	AGTCATTAGC	CATATTAAAA	1320
30	ATAATATAGA	ACAACCTCCT	AATAATAGAT	ATGTGAAAAC	TATAAAACTT	CCATCTTTAA	1380
	AAGTAGGCAC	TAATATAACC	CTATTTTCAT	TATCTAGATT	ATCATCATAT	ATCTTTAGTT	1440
	TAAGCTTTTT	ATTTAAGTAA	ATGTAAAATG	CTGCAATACC	TATAAATCCT	ATAAAACATA	1500
35	AAGATATTAA	AATCTTATTA	TCTAATTGAA	CTTCAAACGT	ATGTACATAT	TTCCGTAAAA	1560
	TAACTACAAA	TAAAAACGAA	CTACCAGTAA	CTGGCCAGAA	AATATTATTT	TTATTTTGTT	1620
	TATCAACATT	TAAATTTTCA	AGTTCCTTCT	CACTAAGTTT	TGCATACCTT	TTGGGAATGA	1680
40	ACCAATTAAT	AAAAGGAAAA	AAGTATACAA	GCCAAGTGCT	TACTAAATCA	ATTAACAAAT	1740
	ACTCATCATT	ATATTGAACG	ACTTTATATC	TCGGATTTTT	ATTAATAACC	TTAATATTAA	1800
45	AAAGCAAAAC	TCACCACGCC	CATTTTCATTG	GATTTATATG	ATTGCTAATA	ATATTTTTAG	1860
	CTTCACTAAC	AGCATTCCCA	ACACTATCCA	TGGATTTTTT	TGTAGTTTTT	TTAACAACAT	1920
	CTATACTATT	ATCGATTTTA	TGCCCTACCC	AGTCTACTTT	ATCTTTTAAT	CCAAAAATAT	1980
50	TATTTTGATA	AATTAAATCT	GTTCCTAATG	CAAATACTGT	ACTCATAGCC	AAACCTGCTA	2040
	AAATCACCCA	TCCTACTGGA	TTACTTCCTA	AAACAAAAGT	CGCTAATCCA	GCTCCAACTG	2100
55	CTGTCCCTGC	AGATCCAGCT	GCAAGCGTgC	ATACCATTAT	GCGACAACGC	CTCTCCAAAT	2160

	CCTTTACCTA GGTATTTTCC GCCTTTTGCA AATTTACTAC CATTTTCTAT AAACACATTA	2280
	CCTGATGTAC GTTTGACTTC CACAAATGAA TTTGGACCTG CTGGGCCTTT CACTCCACCT	2340
5	GCTGTATTGa TAAATACACC GAATTTACTT GcATTTATAC CGTCTTGCTC TAAAAGTGTT	2400
	GACGTAATAT CTAATCCTAT ATCTCTTTTA ATACTGTCTT TATTGTCATT TATATATTTT	2460
	AATATACTTT TCGGGATATC GTCTTCTGGA TGTTCCTTGG CATATGCCTT TATAACAGCA	2520
10	AAGTCTGCTT TATTTAAAGT TTCTTCTCT GCTTTATGTT CAATTTTCCC CATAGCAACT	2580
	TTCAAATATT TTTCATGACT TGCTTTGGCC CAATCAAGTT CTTTACCTGA AGGAATATTA	2640
15	AATTGATTTG TTGAAAAGTT CCAAAAATTC TGCGCTTGGG TAAGTCCTG TTGGACAATT	2700
	TTTTGAAATT CTTCAACTTC TTTAAATATT TCTGGTGATT TTTGATTAAA CTCACGCAAT	2760
	TTGCGTAGCT TCTCTTCTAA TTCATGTTTT TGTTGACCTA ATGTTCTGAT TATTTGTTGG	2820
20	TTGATGAAA TGGCTTGCTG ATTATCGGAA GCATGCTTTT TCAAATTGTT ATTCAAATTT	2880
	TCATATCGCG TAATTGTGTT ACTTAATGAT CTGATATCTT CTTCAAGCTC TGATTCTTTT	2940
	AAAGATATGC TATCAACCTC ACTCGTATAA CGTGACACAA AATTaTCGCA AGCTTGCTTC	3000
25	GTTAAATCAC TCAATGTTTT CATACTTGTT GATAATGGAA TTAACACCGT ACTAAAAAAT	3060
	TGCTTAGCTG ACGTATACGC TTTCCCTTTA AGCGCATCAT CATTAAATAA TTGAGTAATT	3120
30	GCTTTTTCCA ACGCATCATA ATTTGAATTC ATTGTTTGAC TCAAATTCCC CACACTTGAA	3180
	GCTTGGTTTT GAGATCTGTC TAAATACATG TCAATACTCA TCGGCATGCT CCTTTTTCAA	3240
	AAATATATGA TTTTCAAACCT ATTTAAAATC AAATGCTTTT TACATCTACA AAGTTGTAAA	3300
35	ATTTTAAAC TCGGCGATGA TTATTTCTTA TGTAAGGAG TCTAGATGCA GGTAAATTGA	3360
	GATAACATGT CGCCTTTTTT CTTATTTTAG CATATGGATA TAATGGTGTC TTTGTATATT	3420
	CGCAATTAAT CAATAAAAAT TATCTTTCAA TATTTTAATT TTATTGCGAC AACATCCTTA	3480
40	ACATTAAATA TATTAATATC TCAAAATATA TTCACTATTA AAATATGTCA TCAGTTGTTA	3540
	AAAGTATTTT CTCATCATGC GAAATATCAA AACGTATCTA AAATACGAAT AAGTTTATAC	3600
45	AATCACACAA CATCATCATT CAAAATTTTA TTG	3633

(2) INFORMATION FOR SEQ ID NO: 95:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2365 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

	TGATACGAAt GCATTACAAT TCATATGCAA CATACAATTC CTTCTACAGC AAATGAAGTG	60
	AAACAAATAG TTGATGTGAC ATCTGTAGCA GAAAATGATA CGCATTAGTC ATAAAATTAA	120
5	ATGGAAATGT CGATGAAGTG TATCAGCAAT TACAGCGATT AATTAAGAAT GCTAATGTCG	180
	AAGAGAGTGA GAATACTGAC AATATTAATA GTCAAGATAC AAGTTATACA CCTCAAGTAA	240
	AAGTAACAAC ACCAATTTTA GTGAAAGCAC CAATCGCTGG TCGTCGTATT TTACTIONAAAG	300
10	AAGTAAGAGA TTCAATTTTT AGAGAGAAAA TGGTAGGTGA AGGCTTAGCA ATCAAAGCTC	360
	ATGAAGAATC CAAAGTAATC GCACCGTTCA ATGGTTTAAT ATCTATGATT GTACCAACTA	420
15	AGCATGCAGT TGGTATTCAA TCAGAAGACG GTGTGGACAT AGTCATTCAAT ATTGGCGTGA	480
	ATACAGTTGA CTTGGAAGGT AAAGGGTTCA AGTGCTTTGT AAAGCAAAAT GATCATGTTG	540
	AAGCAGGGCA AACGTTGTTG CAATTCGACC AGCAATATAT ACAACAACAA GGCTACAATG	600
20	CTGACGTTAT TGTCGTTATT AGCAACTCTG CCGATTTAGG AAAAGTAGAA CTGACAATGA	660
	ATGAAATCAT TACGACTGAA GATGTTATTT TTAATAATATT TAAAACTAG GAGTGTGTTG	720
	TAATAATGAC AAAATTACCG CAAAATTTCA TGTGGGGTGG CGCTCTTGCC GCAAATCAAT	780
25	TTGAAGGTGG ATATGATAAA GGTGGTAAAG GGTAAAGTGT AATTGATGTT ATGACGAGTG	840
	GTGCACATGG CAAAGCACGT CAGATTACAG AATCTATAGA TCCCAATCAC TATTATCCAA	900
	ATCATGAAGG TATTGATTTT TATCATCGTT ATAAGGAAGA TATTGCCTTG TTTAAAGAAA	960
30	TGGGATTGAA ATGTTTACGT ACGTCGATTG CGTGGACACG TATCTTTCCG AATGGGGATG	1020
	AAGATGTGCC AAATGAAGAA GGACTIONCCT TTTATGATCG TATCTTTGAT GAATTAATTG	1080
35	CACAAGGTAT TGAACCTGTT GTGACGTTAT CACATTTTGA GATGCCACTT CATTTAGCGA	1140
	AACATTATGG TGGATTTAGA AATAGAGAAG TTGTCGATTA TTTTGTGCAT TTTGCGCGTG	1200
	TTGTATTTGA AAGATATAAA GATAAAGTTA CATATTGGAT GACGTTAAT GAAATTAATA	1260
40	ATCAGATGGA CACATCAAAT CCTATCTTTT TATGGACGAA TTCTGGGGTA GCATTGACAG	1320
	AAAATGATAA TCCTGAAGAA GTCyTGTATC AAGTAGCACA TCATGAACTT TTAGCCAGTG	1380
	CyTTAGCAGT TCGTCTTGGT AAAGaGATtA ATCCgAaGTT TAAGATTGGr ACmATGATTt	1440
45	CAmaTGTACC CmTTTATCCa TAwTCGTGTC ATCCGAAAGA TATGATGGAA GCACAAATTG	1500
	CGAATCGCTT ACGTTTCTTT TTCCCGGATG TCCAAGTGAG AGGTTATTAT CCAAGCTATG	1560
	CTAAAAAAAT GTTGGCACGA AAAGGATATG ATGTTGGATG GCAAGAAGGG GACGACAGTA	1620
50	TTTTACAGCA GGGCACGGTT GATTATATTG GCTTTAGTTA TTACATGTCT ACGGCTGTAA	1680
	AACATGATGT TGATACTACA GTTGAAAACA ACATCGTCAA CGGTGGTTTG AATCATTCTG	1740

55

5 GATATACATT GAATGTGTTA TATGATCGTT ATCAGTTACC ACTTTTTTATT GTGGAAAATG 1860
 GTTTTGGTGC AGTTGATGAA GTGGTAGATG GACATATTCa TGATGATTAT CGCATTGAAT 1920
 10 ATTTAAAAGC ACATATTACA GCAGCGATAG AAGCAGTTGA TCAAGATGGT GTAGATTTAA 1980
 TCGGTTATAC ACCGTGGGGA ATCATTGATA TTGTTTCATT TACAACCGGT GAAATGAAGA 2040
 AACGCTATGG TTTAATATAT GTTGATCGAG ATAATGATGG TCATGGCACG ATGGAACGCT 2100
 15 TGAAAAAAGA TTCGTTCTAT TGGTATCAAC AAGTGATAGC ATCAAATGGA GATAAATTAT 2160
 AAAGGTATAT TATAAGTATT TTAGGGTTAG AGCCCCGAGAC ATAAATTAAT ATAGTAGGAC 2220
 CTACAGTGTT ATAATGGCGG gCCCCCAACA CAAAGAATTT CGAAAAGAAA TTCTAcAGGT 2280
 aATGCaAGtT GGCGGGGcCC AACACAGAGA AATTCGAAAA GAAATTCTAc AGGTAATGCA 2340
 AGTTGGGGAA GGACAGAAAT AAATT 2365

20 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11050 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

CTGCGATACG ATTTGTTGAA AGTGGGGAAA ACAAAAAAGT TATCATTACC AATTTAGAGC 60
 AGGCATACGA AGCTTTGATT GGTAATAAAG GTACACACAT TCACATGTAG CACTTTATCA 120
 35 CGCGACAAAA CATTAAATAT GTTTCTCCGT TGATTCAAAT GAAAAAGTTG TCTGCTGACA 180
 CTTTGCAAGG TTTGAAGGAG TTAACTTAT GACAGAAAAC TTTATTTTGG GTAGAAATAA 240
 TAAATTAGAA CATGAACATA AGGCATTAGC AGATTACATT AATATAACCAT ATAGTATATT 300
 40 ACAACCATAT CAAAGTGAAT GTTTTGTGCG ACATTATACG AAAGGCCAAG TTATTTATTT 360
 TTCGCCACAA GAAAGTAGCA ATATTTACTT TTTAATTGAA GGTAACATTA TTAGAGAACA 420
 TTACAATCAA AATGGAGATG TATATCGTTA TTTTAATAAA GAGCAAGTAT TATTTCCAAT 480
 45 CAGTAACTTA TTTCATCCGA AAGAGGTTAA CGAATTGTGT ACAGCATTAA CCGATTGTAC 540
 AGTTCTTGGA TTGCCTAGAG AATTGATGGC CTTTTTGTGC AAAGCTAATG ATGATATATT 600
 TTTGACACTT TTTGCATTAA TAAATGATAA TGAGCAGCAA CACATGAACT ATAACATGGC 660
 50 ATTAACAAGT AAATTTGCTA AAGATCGAAT TATCAAATTG ATATGCCATC TATGTCAGAC 720
 AGTAGGATAC GATCAAGATG AATTTTATGA AATCAAACAG TTTTAACTA TTCAACTCAT 780

55

	TGAAAACTT GTTGTTAAAG ATCATAAAAA TTGGTTAGTA AGCAAACATT TATTCAATGA	900
	TGTATGTGTT TAATATACAA TGTAAAATGA ATAAGTTGAA CATGAGGTCT AACGTACATT	960
5	TATACGTTAG GCCTTTTTTG CTAGCATGAT GAATAATTTA AAATGTTAGT TAAATTTGAT	1020
	TGTTGAAATT ACAGTAAAAT TTAAGGTGAT GAAAAATTTA GAACTTCTAA GTTTTTGAAA	1080
	AGTAAAAAAT TTGTAATAGT GTAAAAATAG TATATTGATT TTTGCTAGTT AACAGAAAT	1140
10	TTTAAGTTAT ATAAATAGGA AGAAAACAAA TTTTACGTAA TTTTTTTCGA AAAGCAATTG	1200
	ATATAATTCT TATTTTCATTA TACAATTTAG ACTAATCTAG AAATTGAAAT GGAGTAATAT	1260
	TTTTGAAAAA AAGAATTGAT TATTTGTCGA ATAAGCAGAA TAAGTATTCG ATTAGACGTT	1320
15	TTACAGTAGG TACCACATCA GTAATAGTAG GGGCAACTAT ACTATTTGGG ATAGGCAATC	1380
	ATCAAGCACA AGCTTCAGAA CAATCGAACG ATACAACGCA ATCTTCGAAA AATAATGCAA	1440
20	GTGCAGATTC CGAAAAAAC AATATGATAG AAACACCTCA ATTAAATACA ACGGCTAATG	1500
	ATACATCTGA TATTAGTGCA AACACAAACA GTGCGAATGT AGATAGCACA ACAAACCAA	1560
	TGTCTACACA AACGAGCAAT ACCACTACAA CAGAGCCAGC TTCAACAAAT GAAACACCTC	1620
25	AACCGACGGC AATTAAAAAT CAAGCAACTG CTGCAAAAAT GCAAGATCAA ACTGTTCTC	1680
	AAGAAGCAAA TTCTCAAGTA GATAATAAAA CAACGAATGA TGCTAATAGC ATAGCAACAA	1740
	ACAGTGAGCT TAAAAATTCT CAAACATTAG ATTTACCACA ATCATCACCA CAAACGATTT	1800
30	CCAATGCGCA AGGAACTAGT AAACCAAGTG TTAGAACGAG AGCTGTACGT AGTTTAGCTG	1860
	TTGCTGAACC GGTAGTAAAT GCTGCTGATG CTAAAGGTAC AAATGTAAAT GATAAAGTTA	1920
	CGGCAAGTAA TTTCAAGTTA GAAAAGACTA CATTTGACCC TAATCAAAGT GGTAACACAT	1980
35	TTATGGCGGC AAATTTTACA GTGACAGATA AAGTGAAATC AGGGGATTAT TTTACAGCGA	2040
	aGTTACCAGA TAGTTTAACT GGTAATGGAG ACGTGGATTA TTCTAATTCA AATAATACGA	2100
40	TGCCAATTGC AGACATTAAA AGTACGAATG GCGATGTTGT AGCTAAAGCA ACATATGATA	2160
	TCTTGACTAA GACGTATACA TTTGTCTTTA CAGATTATGT AAATAATAAA GAAAATATTA	2220
	ACGGACAATT TTCATTACCT TTATTTACAG ACCGAGCAAA GGCACCTAAA TCAGGAACAT	2280
45	ATGATGCGAA TATTAATATT GCGGATGAAA TGTTTAATAA TAAAATTACT TATAACTATA	2340
	GTTCGCCAAT TGCAGGAATT GATAAACCAA ATGGCGCGAA CATTTCTTCT CAAATTATTG	2400
	GTGTAGATAC AGCTTCAGGT CAAAACACAT ACAAGCAAAC AGTATTTGTT AACCCTAAGC	2460
50	AACGAGTTTT AGGTAATACG TGGGTGTATA TTAAAGGCTA CCAAGATAAA ATCGAAGAAA	2520
	GTAGCGGTAA AGTAAGTGCT ACAGATACAA AACTGAGAAT TTTTGAAGTG AATGATACAT	2580

55

	ACCAATTTAA AAATAGAATC TATTATGAGC ATCCAAATGT AGCTAGTATT AAATTTGGTG	2700
	ATATTACTAA AACATATGTA GTATTAGTAG AAGGGCATT AAGGCAATACA GGTAAGAACT	2760
5	TAAAACTCA GGTTATTCAA GAAAATGTTG ATCCTGTAAC AAATAGAGAC TACAGTATTT	2820
	TCGGTTGGAA TAATGAGAAT GTTGACGTT ATGGTGGTGG AAGTGCTGAT GGTGATTGAG	2880
	CAGTAAATCC GAAAGACCCA ACTCCAGGGC CGCCGGTTGA CCCAGAACCA AGTCCAGACC	2940
10	CAGAACCAGA ACCAACGCCA GATCCAGAAC CAAGTCCAGA CCCAGAACCG GAACCAAGCC	3000
	CAGACCCGGA TCCGGATTTCG GATTGAGACA GTGACTCAGG CTCAGACAGC GACTCAGGTT	3060
15	CAGATAGCGA CTCAGAATCA GATAGCGATT CGGATTGAGA CAGTGATTCA GATTGAGACA	3120
	GCGACTCAGA ATCAGATAGC GACTCAGAAT CAGATAGTGA GTCAGATTCA GACAGTGACT	3180
	CGGACTCAGA CAGTGATTCA GACTCAGATA GCGATTGAGA CTCAGATAGC GATTGAGACT	3240
20	CAGACAGCGA TTCAGATTCA GACAGCGACT CAGATTGAGA CAGCGACTCA GACTCAGATA	3300
	GCGACTCAGA CTCAGACAGC GACTCAGATT CAGATAGCGA TTCAGACTCA GACAGCGACT	3360
	CAGACTCAGA CAGCGACTCA GACTCAGATA GCGACTCAGA TTCAGATAGC GATTGAGACT	3420
25	CAGACAGCGA CTCAGATTCA GATAGCGATT CGGACTCAGA CAGCGATTCA GATTGAGACA	3480
	GCGACTCAGA CTCGGATAGC GATTGAGATT CAGATAGCGA TTCGGATTCA GACAGTGATT	3540
	CAGATTGAGA CAGCGACTCA GACTCGGATA GCGACTCAGA CTCAGACAGC GATTGAGACT	3600
30	CAGATAGCGA CTCAGACTCG GATAGCGACT CGGATTGAGA TAGCGACTCA GACTCAGATA	3660
	GTGACTCCGA TTCAAGAGTT ACACCACCAA ATAATGAACA GAAAGCACCA TCAAAATCCTA	3720
35	AAGGTGAAGT AAACCATTTCT AATAAGGTAT CAAAACAACA CAAACTGAT GCTTTACCAG	3780
	AAACAGGAGA TAAGAGCGAA AACACAAATG CAACTTTATT TGGTGCAATG ATGGCATTAT	3840
	TAGGATCATT ACTATTGTTT AGAAAACGCA AGCAAGATCA TAAAGAAAAA GCGTAAATAC	3900
40	TTTTTTAGGC CGAATACATT TGTATTCGGT TTTTTTGTG AAAATGATTT TAAAGTGAAT	3960
	TGATTAAGCG TAAAATGTTG ATAAAGTAGA ATTAGAAAGG GGTCATGACG TATGGCTTAT	4020
	ATTTTCATTAA ACTATCATT CACCAACAATT GGTATGCATC AAAATTTGAC AGTCATTTTA	4080
45	CCGGAAGATC AAAGCTTCTT TAATAGCGAT ACAACTGTTA AACCATTAAA AACTTTAATG	4140
	TTGTTACATG GATTATCAAG TGATGAAACG ACATATATGA GATATACAAG CATAGAAAGG	4200
	TATGCGAATG AACACAAATT AGCTGTGATT ATGCCCAATG TGGATCATAG CGCATATGCT	4260
50	AACATGGCAT ATGGTCATAG CTATTATGAT TATATTTTGG AAGTGTATGA TTATGTTTAT	4320
	CAAATATTTT CACTTTCCAA AAAGCGTGAT GACAATTTTA TAGCAGGTCA CTCTATGGGA	4380

55

	TTATCTGCTG TGTTTGAAGC GCAAAATTTA ATGGATCTAG AGTGGAATGA TTTTTCAAAA	4500
	GAGGCCATAA TTGGCAATCT TTCAAGTGTT AAAGGAACTG AACATGATCC GTATTACTTG	4560
5	CTAGACAAAG CTGTAGCTGA AGATAAACAA ATTCCAAAAT TGCTCATTAT GTGTGGTAAA	4620
	CAAGACTTTT TATATCAAGA CAACTTAGAT TTTATCGATT ATTTATCACG CATAAATGTT	4680
	CCTTATCAAT TTGAAGATGG ACCAGGAGAT CATGATTATG CATATTGGGA TCAAGCGATT	4740
10	AAGCGTGCTA TAACATGGAT GGTGAATGAT TAATTATTTT TTGGAAAATA TGTGGCTGCA	4800
	TTAAATACAC AGAGTGAGAG ATACAACTA TTTACGCACG ACTAACATTT CTAAGTGTTC	4860
15	AAATTATTTT TGTATTAATA TGATTGGCGC AATTTGCTGA TACACAAAAA TGTTCCTCGT	4920
	GAAACTTAGA TTTAGCTTAT AGTTTTATCA TCATTTGTAT GACTTACATT ATAAATTTTA	4980
	TTATAATGAG GTTAAACGCTT TGAAAGGAGT CATCATCATG TCGACCAATA AAAACGATTA	5040
20	TGAGCATATG TTGTTTTATT TTGCATATAA AACCTTTATT ACTACCGCTG ATGAAATTAT	5100
	AGAGAAGTAT GGTATGAGTC GTCAGCATCA TCGTTTTTTG TTTTTTATCA ATAAATTACC	5160
	TGGTATTACT ATTAAATCAT TACTAGAAAT ATTAGAAATT TCTAAmCAAG GATCACATGC	5220
25	AACACTTCAA AAATTAAAAG AGCAAGGTCT CATTATTGAA AAAGTTTTAG AGACTGATCG	5280
	ACGTGTCAAA AAATTATATT CGACGGATAA AGGCGATCAA CTCATTGCTG AATTGAACAA	5340
	GGCGCAAGAT GAATTATTGC AAAATATATA TCAACAAGTC GGTTCGGATT GGTATGATGT	5400
30	GATGGAAGCA TTGGCTAAAG GgCGACCTGG cTTTGATTTT ATTAAGCATT TGAAAGATGA	5460
	AAAAGAAAGC TAGCATCAGA AATGTTAAAA ATCTTCGCAT TCTTAAATTT AAAAAATATG	5520
35	TCAAAAAGTG TATAATAAAA ACATATAATT TAATTGAACT CAGTTTCAAC ACATCTTAGA	5580
	AAGGAGTTTG AATGATGAAA AAATTAGCAG TTATTTTAAC ATTAGTTGGC GGTTTTACT	5640
	TCGCATTTAA AAAATACCAA GAACGTGTTA ACCAAGCACC TAACATTGAG TACTAAATTA	5700
40	AACCATAAAA AATTCCCGAA CACCTTGTTA TAGTGCTCGG GAATTTTTTT ATGCTTTACT	5760
	TGAATATATC AAATATTATT TTTGCGCTTT CTGTATTTTC GATAATTACCA CTAAATGATT	5820
	CTGATCTAGG TCCGTAAGCG TAGGTATTAA CATCCTCGCC TGTATGTCCA TCGGAAGTCC	5880
45	ACCCTGTATA AGATTTATCA TTTACTGGCT TCTGAATAGC GTGTTGTAGG GCTTTTGTTT	5940
	GCGTTTCTAC TTCTGCGGAT TTTTCGTCTT TTTCTTTTTT AAGTAGTCTT TTTAGCTTTT	6000
	TATTCTCTTT TTTAACCTTT TTCATATCAT CTTGTGAAAA TTCAAATCCA TAACCTTCAT	6060
50	TAATAACTTT TTCAGGGTCT TCACCTTTAG CCATTTTTTC TGTCATATAT GATCCAGAGT	6120
	GTTTCATAGA TTTAATCGGT TGAGGATTCC ATTCGTATCC TTTATCTTTA CCAATTGTTA	6180

55

	ATTGAATGGC GTCATCGAAT GCTTTTTTCAA AACCTTCCAT TTCAGACATA ACGCCTGTAA	6300
	TATCGTTGGA ATGCGCTGAT TTATCTATAG AAGCACCTTC GACCATTAAA AAGAATCCTT	6360
5	TTTTATTGCG CTCAAGCTTA CTAAGTGCAC TTTGTTGCAT ATCAGCTAAT GATGGTTCGT	6420
	CTTTAGAAGC ATCTATTGCA AGTGGCATAT TTTTATCTGC AAACAAACCA AGAACTTTAT	6480
	CTTTATCAGA TTTTGATAAC TCCTTACTGT TCGTGGCAAG GTCGTAACCA TCTTTTTTGA	6540
10	ATTTTTTATC TAAATTGCCA TTACTTTTAC CGAAATATTT AGCGCCGCCG CCTAATAAAA	6600
	CATCAACTTT ATGCTTTCCG TTGATTTTAT CTTTATAAAA TTGTTTAGCG ATTTCTGTTTT	6660
	TATCATCTCT AGAAGTCACG TGTGCAGCAT ATGCTGCTGG TGTTCATCT GTTAATTCAG	6720
15	CTGTTGAAAC AAGACCAGTC GACTTACCTT TTTCTTTTGC ACGTTCAAGC ACCGTCTTTA	6780
	CTTCTGCTT GTTACTGTCA ACACCGATGG CACCATTATA TGTCTTATGA CCAGAACTAA	6840
20	AGGCTGTTCC GCCAGCTGCA GAATCAGTAA TATTCTGTTT TGGGTCATTT GAATATGTAC	6900
	GATTTGTGCC TTTTAAATAT GAATCAAAAG CAGTAGGGGT CATTTCTTTA GCATGCGGAT	6960
	CATTTTATA ATAACGATAA GCTGTGTTAA ATGATGGACC CATGCCATCG CCAACTAAAA	7020
25	AGATAACATT TTTTGGATTT TTAGTATTAC CAACCGCGAA ACTTTCATCT TTAGAACTTT	7080
	TATCGGATTG CGCAATTGCA GGTGTGACAG AACTAAAAAC CGTTGACACG ATAATAAGGT	7140
	TAGCAACTGC AAATTTTGTG GCTTTTTTAA CTGATAACAT AAGACATCCT CCTGAGTATA	7200
30	TGACTATGTC TTCAGTGTA AAGAGGAATT TtGAGCAATT ATGTAGTTTT AGTTAnAAAT	7260
	ATGTAACAG AGTGATTAG AATAACAAAA aATGAATATA TATGACAATT TGTTATAGAA	7320
	AGCGTTAGAA TAGAAGCGTG TGAAAATATA GAATTAAATA TAATTTGAGG TGGAAAAATG	7380
35	ATACTAGTAA TGTTATCTCC ATTATTAATC ATATTCTTTA TAGTGTGTC TATTTTAGAA	7440
	GAGCGTAAAC GTACGAAGAA AAAGCAACTC GAGAAAGAAA AAGCAAATAC ACTAAATCAA	7500
40	AATACAAATG ACACGGAAAG TTCAAATCAA GAGCCGTCAT TGCAGCAGGA TAAAGAACAA	7560
	AAAGATAACA AAGGATAATT CAATTGAAGG AAGAAGATTA TAGATGAAAA TATTAATTGT	7620
	TGAAGATGAT TTTGTTATAG CAGAGAGTTT AGCATCTGAA CTTAAAAAAT GGAATTACGG	7680
45	TGTTATTGTC GTTGAACAAT TTGATGATAT ACTGTCTATC TTAAACCAA ATCAACCTCA	7740
	GCTTGTATTG CTAGATATTA ATTTGCCAAC GTTAAATGGT TTTCATTGGT GTCAAGAAAT	7800
	CCGAAAAACA TCTAATGTGC CAATTATATT TATTAGTTCC CGTATTGATA ATATGGACCA	7860
50	AATTATGGCA ATACAAATGG GGGGAGATGA TTTTATCGAA AAGCCATTTA ACTTGTCATT	7920
	AACGATTGCC AAAATTCAAG CATTATTGAG ACGAACTTAT GACTTGTGAG TAGCTAATGA	7980

55

	ACAAAACATA CAGCTATCTT TGA CTGAATT ACAAATATTA AAGTTATTAT TTCAAAATGA	8100
	AGaTAAATAT GTAAGTAGrA CTGCTTTAAT TGaAAAATGT TGGGaATCAG AAAA CtTCAT	8160
5	AGATGATAAC ACATTAGCTG TTAACATGAC GCGCCTGCTG AAAAAATTAA ATACTATTGG	8220
	CGTTAATGAT TTTATCATTa CAAAGAAAAA TGTCGGATAT AAAGTATAGG GTGAATGCAA	8280
	TGACCTTTCT TAAAAGTATT ACTCAGGAAA TAGCAATAGT CATAGTTATT TTTGCTTTGT	8340
10	TTGGCTTAAT GTTTTACCTG TATCATTGTC CATTAGAAGC ATATTTACTA GCACTTGGCG	8400
	TTATTTTATT ATTATTACTC ATATTCATAG GTATTAAATA TTTAAGTTTT GTAAAACTA	8460
15	TAAGCCAACA ACAACAAATT GAAAACTTAG AAAATGCGTT GTATCAGCTT AAAAATGAAC	8520
	AAATTGAATA TAAAAATGAT GTAGAGAGCT ACTTTTTTAAC ATGGGTACAT CAAATGAAAA	8580
	CACCCATTAC TGCAGCACAA CTGTTACTTG AAAGAGATGA GCCTAATGTT GTTAATCGTG	8640
20	TTCGTCAAGA GGTATTTCaA ATTGaTAACT ATACAAGTTT AGCACTTAGT TATTTAAAGT	8700
	TATTAAATGA AACTTCTGaT ATTTCTGTCA CTAAATTTTC GATTAAATAAT ATCATTGCCC	8760
	CAATTATTAT GAAATATTCA ATACAGTTTA TTGATCAAAA AACAAAAATC CATTATGAAC	8820
25	CTTGTCATCA CGAAGTATTA ACTGACGTTA GATGGACCTC TTTAATGATA GAACAATTAA	8880
	TAAATAATGC ACTTAAGTAT GCGAGAGGTA AAGATATATG GATTGAATTT GATGAGCAAT	8940
	CCAATCAATT ACACGTAAAA GATAATGGTA TCGGTATTAG TGAAGCGrAC TTGCCTAAAA	9000
30	TATTTGATAA GGGCTATTCA GGTATAATG GCCAGCGCCA AAGTAACTCA AGTGGGaTTG	9060
	GTTTATTTAT CGTAAAACAA ATTTCAACAC ACACAAACCA TCCTGTTTCA GTCGTATCTA	9120
35	AACAAATGA GGTACAACA TTTACGATTC AATTTCCAGA TGAATAAAAA CTTTCAATAT	9180
	TGTAAGTATA CTAGTAACAT TTTTTTACTA ATTTAAATGT TATTAGTATT TTTTTGTTTT	9240
	AATATAGAAC TAACAAAGAA ATGAGGTGCA TGCCATGTTG CTAGAAGTGn AACATGTAAA	9300
40	AAAGGTTTAT GGTAAAGGTT TGAATGCTAC GACAGCACTT AATCAAATGA ATTTATCAGT	9360
	TGGAGCTGGT GaATTTGTTG CaATTATGGG TGAGTCTGGG tCAGGGAAAGT CTACACTACT	9420
	AAATTTAATT GCTCTTTTTG ATGGACTAAC TGAAGGTGAC ATTATTGTGG ATGGCGCACA	9480
45	TTTAAATAAT ATGAAAAATA AAAGTAAAGC ATTGTATCGT CaACAAATGG TAGGTTTTGT	9540
	TTTTcAAGAT TTTAATCTTT TACCAACAAT GACGAATAAA GAAATATAA TGATGCCATT	9600
50	AATTTTAGCT GGTGCTAAAC GAAAAGATAT AGAACAAAGG GTACATCAGT TGGCAGTACA	9660
	ATTACATTTA GAGGGATTCT TAAACAAGTA TCCTTCTGAA ATCTCTGGGG GTCAGAAGCA	9720
	ACGCATTGCC ATTGCACGTG CATTAGTTAC TAAGCCGACG ATTTTACTAG CCGATGAACC	9780

55

TCAATTGGAA CAGACAATTT TAATGGTAAC TCATTCAAAT ATCGATGCGT CTTATGCAGA 9900
 GCGAGTCAATT TTTATTAAAG ATGGGCGTCT ATATCATGAA ATATATCGTG GTGAAGAAAG 9960
 5 TCAATTAGCT TTTCAACAAC GAATAACAGA TAGCTTAGCA CTTGTGAATG GAGGAAGTGT 10020
 CAATATATGA AGTTAAGATT GTTATGnACA TAGTGCGACG TCAATTTATT ACGCAGCGAC 10080
 TTGTAATCAT TCCATTCATT TTAGCGGTAA GTGTACTATT CATGATTGAA TATACGCTTG 10140
 10 TGTCAATTGG GTTAAATAGC TACATAAAAC AGAAGAATGA CTTCTTAGTA CCATTTATTA 10200
 TCATAGCTAA TTTTTTTATG GCGCTTTTAA CTTTTATTTT TATTTTCTAT GCAAATCACT 10260
 TTATGATGTC ACAAAGACGA AAAGAGTTTA GCATTTTTTAT GACATTGGGC ATGACCAAGA 10320
 15 AAAGTATGCG TTTAATTGTA GTGATGGAAA CTATCTTACA ATTTGTGATA ATTCAGTCG 10380
 TTAGTATTGC CGGCGGATAC TTAAGGTATT ATCCATTTGA CTCTGTAGCG ATGTTTATTA 10440
 20 TGGGCAGTGA AGTTGCGACG TTAAGGTATT ATCCATTTGA CTCTGTAGCG ATGTTTATTA 10500
 CTTTGATTAT CATTGCTGTA TTAATGGGCA TGCTACTTAT ATTCAACTTG TTTAGTATTA 10560
 ATTTTCAACG GCCGATAACT TATCAACATC GTTCCGATTC TAGTGTGATA TCACGATGGT 10620
 25 TGCCTTACGT TTTAATTGTT ATAGGAAGCG CAACTATAT TTAGGTTACT TTATTGCATT 10680
 ACAACAAGAT ACGACGTTTG GTGCCTTTTT TAAAATATGG ATTGTCATAG GATTAGTTAT 10740
 TATCGGTACT TATGCATTTT TTGTAGGTAT AAGTGAAATA ATTATTAGTA TATTGCAGCA 10800
 30 GGTATCAAAA GTTTACTATC ATCCACGGTA TTTTTTTGTG GTAGTTGGGA TGCCTGTACG 10860
 TCTTAAAATG AATGCAGTCA GTCTTGCAAC AATCACTTTG CTGTGTACAT TTTTGATTGT 10920
 AACGCTCACA ATGACATTAA CAACCTATCG TGATATGAAT CATACCATTA CGAAATTGAT 10980
 35 TACGAATGAT TakGATTTGT CATTTAGCGA CAATTCTAAG TCACAAaTAG AACGTCAACA 11040
 AACaATTGAG 11050

40 (2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 983 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 45 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

50 CGACATAACG AGGCAAGGGT ACATGATACT TTAGCCTCGT TTTTGATATG TATTTTTCTG 60
 AATATAAGGG CAATAGATGG TATTTTATAw TTTTTTTAAG GTAGTGATTA ACATAGATAT 120

55

TCAAGCGGAA CAGCATTATG CACCAGTATT AACGCATTTT TTAGATCCAA GAGGGCAATA 240
 TATATTGGAA GTGATTGTG GCAGTTATGA AGATTTAAAC GTATCTTTT ATGGTGGACC 300
 5 TAATGCTGAA AGAAAAAGAG CAATCATTTT GCCGAACAT TATGAACCTA AAGAAAGCGA 360
 CTTTGAATTA ACTTTAATGG AAATAGATTA TCCTGAAAAA TTCGTCACTT TAAAACATCA 420
 ACATATTTTA GGGACATTAA TGTCTTTAGG TATCGAACGC GAACAAGTTG GAGATATAAT 480
 10 TGTGaatGAA CGAATTCAAT TTGTTTTGAC AAGTAGATTG GAATCATTTA TTATGTTAGA 540
 ATTACAACGT ATTAAAGGCG CATCAGTTAA ACTTTATACT ATTCCAGTAA CAGATATGAT 600
 ACAATCTAAT GAGAATTGGA AAAATGAAAG TGCaCAGTTA GTTCTTTAAG GTTAGATGTT 660
 15 GTTATTAAAG AAATGATACG TAAATCACGT ACGATTGCGA AACAACTAAT CGAAAAAAAA 720
 CGTGTTAAAG TGAATCACAC TATTGTTGAT TCAGCAGATT TTCAATTACA AGCAAATGAT 780
 20 TTAATATCCA TCCAAGGTTT TGGTAGAGCA CACATTACTG ACTTAGGTGG TAAACTAAA 840
 AAAGATAAAA CGCACATTAC CTATAGAACA TTATTCAAAT AGTAATGATT TAAGGAGGAT 900
 AACAAATGCC TTTTACACCA AATGAaATTA AGAATAAAGA GTTTTCACGT GTaAAGAATG 960
 25 GTTTTAGAAC CTACTIONAGT TGG 983

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
 30 (A) LENGTH: 10322 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

TTTTGCAAAG CTTATTTTAT GTCAAACAGA TAGTCAATGT GAAACAAAGG TTAGTACATA 60
 40 TAATCATCCA GACTTTTATGT ATATATCAAC AACTGAGAAT GCAATTAAGA AAGAACAAGT 120
 TGAACAACTT GTGCGTCATA TGAATCAACT TCCTATAGAA AGCACAAATA AAGTGACAT 180
 CATTGAAGAC TTTGAAAAGT TAACTGTTCA AGGGGAAAAC AGTATCTTGA AATTTCTTGA 240
 45 AGAACCACCG GACAATACGA TTGCTATTTT ATTGTCTACA AAACCTGAGC AAATTTTAGA 300
 CACAATCCAT TCAAGGTGTC AGCATGTATA TTTCAAGCCT ATTGATAAAG AAAAGTTTAT 360
 AAATAGATTA GTTGAACAAA ACATGTCTAA GCCAGTAGCT GAAATGATTA GTACTTATAC 420
 50 TACGCAAATA GATAATGCAA TGGCTTTAAA TGAAGAATTT GATTTATTAG CATTAAGGAA 480
 ATCAGTTATA CGTTGGTGTG AATTGTTGCT TACTAATAAG CCAATGGCAC TTATAGGTAT 540

55

	GAATGGTTTC TTCGAAGATA TCATACATAC AAAGGTAAAT GTAGAGGATA AACAAATATA	660
	TAGTGATTTA AAAAATGATA TTGATCAATA TGCGCAAAAG TTGTCGTTTA ATCAATTAAT	720
5	TTTGATGTTT GATCAACTGA CGGAAGCACA TAAGAAATTG AmTCAAAATG TAAATCCAAC	780
	GCTTGTATTT GAACAAATCG TAATTAAGGG TGTGAGTTAG ATGCCAAATG TAATAGGTGT	840
10	TCAGTTTCAA AAAGCGGGAA AATTAGAATA TTATACACCT AATGATATAC AAGTAGATAT	900
	AGAAGACTGG GTAGTTGTCG AATCTAAAAG AGGCATAGAG ATAGGTATTG TTAAAAATCC	960
	ATTAATGGAT ATTGCTGAAG AGGATGTTGT GTTACCTCTT AAAAATATTA TTCGCATTGC	1020
15	TGATGACAAA GATATTGATA AATTTAATTG TAATGAACGA GATGCTGAAA ATGCATTAAT	1080
	ACTATGTAAA GACATTGTAA GAGAACAAGG TTTGGACATG CGTTTAGTCA ATTGCGAATA	1140
	TACATTAGAT AAATCGAAAG TTATTTTTAA TTTTACGGCG GATGATCGTA TTGATTTTAG	1200
20	AAAATTAGTA AAAATATTAG CGCAACATTT AAAACACGT ATCGAGTTGA GACAAATTGG	1260
	TGTAAGGGAT GAAGCCAAAT TGCTTGGCGG TATCGGACCT TGTGGTAGGT CGTTATGTTG	1320
	TTCTACATTT TTAGGGGATT TTGAACCAGT ATCGATTAAG ATGGCTAAGG ATCAAAATTT	1380
25	ATCATTAAT CCAACTAAAA TTTCTGGTGC ATGTGGTCGT TTGATGTGTT GTTTAAAATA	1440
	TGAAAATGAC TATTATGAGG AAGTACGTGC ACAATTACCT GATATTGGTG AAGCAATTGA	1500
30	AACGCCTGAT GGTAACGGGA AAGTAGTTGC TTTAAATATA TTAGACATTT CTATGCAGGT	1560
	GAAGCTTGAG GGACATGAAC AGCCACTTGA ATATAAATTA GAAGAAATAG AAACATGCA	1620
	TTAAGGAGGC ATTATTACAT TTGGATCGCA ATGAAATATT TGAAAAAATA ATGCGTTTAG	1680
35	AAATGAATGT CAATCAACTT TCAAAGGAAA CTTCAGAATT AAAGGCACTT GCAGTTGAAT	1740
	TAGTAGAAGA AAATGTAGCG CTTCAACTTG AAAATGATAA TTTGAAAAAG GTGTTGGGCA	1800
	ATGATGAACC AACTACTATT GATACTGCGA ATTCAAAACC AGCAAAAGCT GTGAAAAAGC	1860
40	CATTACCAAG TAAAGATAAT TTGGCTATAT TGTATGGAGA AGGATTTTCAT ATTTGTAAAG	1920
	GCGAATTATT TGAAAAACAT CGACATGGTG AAGATTGTCT GTTCTGTTTA GAAGTTTTAA	1980
	GTGATTAATC AAGCACACTC AAATAGTGTT ATAATTATAA ATGAATATGG TTTGGATAAG	2040
45	TCTGAGACAA TGCAATGTTT AGGCTTTAAT TGTGTATAAA GTTTTGGTGA TTGCATAAGA	2100
	GATGGCGGTA CTAAATGTTA TTATTAAGTG TGCACGCagT ATCaTTAGTT ATAAATGTGA	2160
50	GCTGTAAAA GTCAAAAATA CATCGAATGT AGTTAGGCAT ATAATATAAA AAGAGTTTTTC	2220
	AATTACTCAA TAGAAAAAGG TTGTCTTCAT AGGAGTTAAA AATGTTAAAA GAGAATGAAC	2280
	GATTTGATCA ACTAATCAAA GAAGATTTTA GTATTATTCA AAATGATGAT GTTTTTTCAT	2340

55

	TGGACTTATG TTCAGGCAAT GGGGTGATAC CCTTGTTATT GTTTGCGAAA CATCCACGAC	2460
	ATATAGAAGG TGTGAGATT CAAAAACAC TTGTCGATAT GCGCGACGC ACATTTCAAT	2520
5	TCAATGATGT TGATGAATAT TTAACAATGC ATCACATGGA TTTGAAAAAC GTTACTAAAG	2580
	TATTTAAACC TTCACAATAT ACTTTAGTAA CGTGTAATCC GCCTTATTTT AAAGAGAATC	2640
10	AGCAACACCA ACATCAAAAA GAAGCACATA AGATAGCGAG ACATGAGATT ATGTGTACAC	2700
	TTGAAGATTG CATGATTGCA GCCCGTCATT TATTTAAAGA AGGTGGCAGG CTAAACATGG	2760
	TACATCGTGC AGAGAGACTA ATGGATGTCT TGTTTGAAAT GAGAAAAGTG AATATTGAAC	2820
15	CTAAGAAAGT CGTTTTTATA TATAGTAAAG TAGGGAAATC AGCACAAACG ATAGTAGTAG	2880
	AAGGTCGAAA AGGTGGAAAT CAAGGTTTAG AAATCATGCC CCCATTTTAT ATTTATAATG	2940
	AAGATGGTAA TTATAGCGAA GAAATGAAGG AAGTATATTA TGGATAGTCA TTTTGTATAT	3000
20	ATTGTAAAT GTAGTGATGG AAGTTTATAT ACAGGATACG CTAAAGACGT TAATGCACGT	3060
	GTTGAAAAAC ATAACCGAGG TCAAGGAGCC AAATATACGA AAGTAAGACG TCCGGTGCAT	3120
	TTAGTTTATC AAGAAATGTA TGAGACAAAG TCTGAAGCAT TGAACCGTGA ATATGAAATT	3180
25	AAACTTATA CCAGACAAAA GAAATTGCGA TTAATTAAGG AGCGATAGTA TGGCTGTATT	3240
	ATATTTAGTG GGCACACCAA TTGGTAATTT AGCAGATATT ACTTATAGAG CAGTTGATGT	3300
30	ATTGAAACGT GTTGATATGA TTGCTTGTGA AGACACTAGA GTAAC TAGTA AACTGTGTAA	3360
	TCATTATGAT ATTCCAATC CATTAAAGTC ATATCACGAA CATAACAAGG ATAAGCAGAC	3420
	TGCTTTTATC ATTGAACAGT TAGAATTAGG TCTTGACGTT GCGCTCGTAT CTGATGCTGG	3480
35	ATTGCCCTTA ATTAGTGATC CTGGATACGA ATTAGTAGTG GCAGCCaGAG AAGCTAATAT	3540
	TAAAGTAGAG ACTGTGCCTG GACCTAATGC TGGGCTGACG GCTTTGATGG CTAGTGGATT	3600
	ACCTTCATAT GTATATACAT TTTTAGGATT TTTGCCACGA AAAGAGAAAG AAAAAAGTGC	3660
40	TGTATTAGAG CAACGTATGC ATGAAAATAG CACATTAATT ATATACGAAT CACCGCATCG	3720
	TGTGACAGAT ACATTAAAA CAATTGCAA GATAGATGCA ACACGACAAG TATCACTAGG	3780
	GCGTGAATTA ACTAAGAAGT TCGAACAAAT TGTAAGTAT GATGTAACAC AATTACAAGC	3840
45	ATTGATTCAG CAAGGCGATG TACCATTGAA AGGCGAATTC GTTATCTTAA TTGAAGGTGC	3900
	TAAAGCGAAC AATGAGATAT CGTGGTTTGA TGATTTATCT ATCAATGAGC ATGTTGATCA	3960
50	TTATATTCAA ACTTCACAGA TGAAACCAAA ACAAGCTATT AAAAAAGTTG CTGAAGAACG	4020
	ACAACTTAAA ACGAATGAAG TATATAATAT TTATCATCAA ATAAGTTAAT CACTTTATCG	4080
	ATTaTATGAA ATTTTAAACG ATTTTATAAA CGCAAGCTGT AATTTTAAAT GGTAAGTTAT	4140

GTTTTTTAAT GTAAAATAAA TACATTGAAA GTAATAAATA CCTTAACATT GAATAAGATG 4260
 AAAATGAGAT GACGAGATAA ATGTTTCGCGT CCGTTGAAAT GCATAGAAAT CTTAGATATT 4320
 5 ATTTGAAGTG AGACATTACG AGGAGGAACA GTTATGGCTA AAGAAACATT TTATATAACA 4380
 ACCCCAATAT ACTATCCTAG TGGGAATTTA CATATAGGAC ATGCATATTC TACAGTGGCT 4440
 10 GGAGATGTTA TTGCAAGATA TAAGAGAATG CAAGGATATG ATGTTGCTA TTTGACTGGA 4500
 ACGGATGAAC ACGGTCAAAA AATTCAAGAA AAAGCTCAAA AAGCTGGTAA GACAGAAATT 4560
 GAATATTTGG ATGAGATGAT TGCTGGAATT AAACAATTGT GGGCTAAGCT TGAAATTTCA 4620
 15 AATGATGATT TTATCAGAAC AACTGAAGAA CGTCATAAAC ATGTCGTTGA GCAAGTGTTC 4680
 GAACGTTTAT TAAAGCAAGG TGATATCTAT TTAGTGAAAT ATGAAGGTTG GTATTCTGTT 4740
 CCGGATGAAA CATACTATAC AGAGTCACAA TTAGTAGACC CACAATACGA AAACGGTAAA 4800
 20 ATTATTGGTG GCAAAAGTCC AGATTCTGGA CACGAAGTTG AACTAGTTAA AGAAGAAAGT 4860
 TATTTCTTTA ATATTAGTAA ATATACAGAC CGTTTATTAG AGTTCTATGA CCAAAATCCA 4920
 GATTTTATAC AACCACCATC AAGAAAAAAT GAAATGATTA ACAACTTCAT TAAACCAGGA 4980
 25 CTTGCTGATT TAGCTGTTTC TCGTACATCA TTTAACTGGG GTGTCCATGT TCCGTCTAAT 5040
 CCAAAACATG TTGTTTATGT TTGGATTGAT GCGTTAGTTA ACTATATTTTC AGCATTAGGC 5100
 TATTTATCAG ATGATGAGTC ACTATTTAAC AAATACTGGC CAGCAGATAT TCATTTAATG 5160
 30 GCTAAGGAAA TTGTGCGATT CCACTCAATT ATTTGGCCTA TTTTATTGAT GGCATTAGAC 5220
 TTACCGTTAC CTAAAAAAGT CTTTGCACAT GGTGGAATTT TGATGAAAGA TGGAAAAATG 5280
 35 AGTAAATCTA AAGGTAATGT CGTAGACCCT AATATTTTAA TTGATCGCTA TGGTTTAGAT 5340
 GCTACACGTT ATTATCTAAT GCGTGAATTA CCATTTGGTT CAGATGGCGT ATTTACACCT 5400
 GAAGCATTG TTGAGCGTAC AAATTTTCGAT CTAGCAAATG ACTTAGGTAA CTTAGTAAAC 5460
 40 CGTACGATT CTATGGTTAA TAAGTACTTT GATGGCGAAT TACCAGCGTA TCAAGGTCCA 5520
 CTTCATGAAT TAGATGAAGA AATGGAAGCT ATGGCTTTAG AAACAGTGAA AAGCTACACT 5580
 GAAAGCATGG AAAGTTTGCA ATTTTCTGTG GCATTATCTA CCGTATGGAA GTTTATTAGT 5640
 45 AGAACGAATA AGTATATTGA CGAAACAACG CCTTGGGTAT TAGCTAAGGA CGATAGCCAA 5700
 AAAGATATGT TAGGCAATGT AATGGCTCAC TTAGTTGAAA ATATTCGTTA TGCAGCTGTA 5760
 50 TTATTACGTC CATTCTTAAC ACATGCGCCG AAAGAGATTT TTGAACAATT GAACATTAAC 5820
 AATCCTCAAT TTATGGAATT TAGTAGTTTA GAGCAATATG GTGTGCTTAA TGAGTCAATT 5880
 ATGGTTACTG GGCAACCTAA ACCTATTTTC CCAAGATTGG ATAGCGACGG AnAATTGCAT 5940

55

AACCTCAAAT TGATATTAAA GACTTTGATA AAGTTGAAAT TAAGGCAGCA ACGATTATTG 6060
 ATGCTGAACA TGTTAAGAAG TCAGATAAGC TTTTAAAAAT TCAAGTAGAC TTAGATTCTG 6120
 5 AACAAAGACA AATTGTATCA GGAATTGCCA AATTCTATAC ACCAGATGAT ATTATTGGTA 6180
 AAAAAGTAGC AGTTGTTACT AACCTGAAAC CAGCTAAATT AATGGGACAA AAATCTGAAG 6240
 GTATGATATT ATCTGCTGAA AAAGATGGTG TATTAACCTT AGTAAGTTTA CCAAGTGCAA 6300
 10 TTCCAAATGG TGCAGTGATT AAATAACTGT ATTTTAAAAA ATTAGGAGAG ATAATTATGT 6360
 TAATCGATAC ACATGTCCAT TTAAATGATG AGCAATACGA TGATGATTG AGTGAAGTGA 6420
 15 TTACACGTGc TAGAGAAGCA GGTGTTGATC GTATGTTTGT AGTTGGTTTT AACAAATCGA 6480
 CAATTGAACG CGCGATGAAA TTAATCGATG AGTATGATTT TTTATATGGC ATTATCGGTT 6540
 GGCATCCAGT TGACGCAATT GATTTTACAG AAGAACACTT GGAATGGATT GAATCTTTAG 6600
 20 CTCAGCATCC AAAAGTGATT GGTATTGGTG AAATGGGATT AGATTATCAC TGGGATAAAT 6660
 CTCCTGCAGA TGTTCAAAAG GAAGTTTTTA GAAAGCAAAT TGCTTTAGCT AAGCGTTTGA 6720
 AGTTACCAAT TATCATTCAT AACCGTGAAG CAACTCAAGA CTGTATCGAT ATCTTATTGG 6780
 25 AGGAGCATGC TGAAGAGGTA GCGGGGATTA TGCATAGCTT TAGTGGTTCT CCAGAAATTG 6840
 CAGATATTGT AACTAATAAG CTGAATTTTT ATATTTTCATT AGGTGGACCT GTGACATTTA 6900
 AAAATGCTAA ACAGCCTAAA GAAGTTGCTA AGCATGTGTC AATGGAGCGT TTGCTAGTTG 6960
 30 AAACCGATGC ACCGTATCTT TCGCCACATC CGTATAGAGG GAAGCGAAAT GAACCGGCGA 7020
 GAGTAACTTT AGTAGCTGAA CAAATTGCTG AATTAAAAGG CTTATCTTAT GAAGAAGTGT 7080
 35 GCGAACAAAC AACTAAAAAT GCAGAGAAAT TGTTTAATTT AAATTCATAA AGTTAAAAGT 7140
 GAGAAAGATC ACCGCCATAA ATGTAAACGA TGCTATATTC GTTTAATATG CTATGGTTCT 7200
 TTCTCACTTT TTTAAATTAA AATATCGTGC ATGTGGAATA CGTGCGATAG AGATGGTTAG 7260
 40 AGCTTTGAAA TTAAGAAATTG TAGGAAGGCG TTTTAAATGA AAATCAATGA GTTTATAGTT 7320
 GTAGAAGGAC GAGATGATAC TGAGCGTGTT AAACGAGCTG TTGAATGTGA TACGATTGAA 7380
 ACGAATGGTA GTGCCATCAA CGAACAACT TTAGAAGTAA TTAGAAATGC TCAACAAAGT 7440
 45 CGAGGCGTTA TTGTATTAAC AGATCCAGAT TTCCCAGGAG ATAAAATTAG AAGTACAATT 7500
 ACTGAACATG TCAAAGGTGT TAAACATGCG TATATTGATA GAGAAAAAGC TAAAAATAAA 7560
 AAAGGGAAAA TTGGTGTTGA ACATGCCGAC TTAATTGATA TTAAAGAAGC GTTAATGCAT 7620
 50 GTTAGTTCAC CCTTTGATGA AGCTTATGAA TCAATTGATA AATCTGTGCT AATAGAGTTG 7680
 GGGTTAATTG TTGGGAAAGA TGCAAGGCGC CGTAGAGAAA TTTTAAGTAG AAAATTGCGA 7740

55

	GCGGATGTAA GGCAAGCTTT AGAAGATGAA TGAGGAAGTG AAAATGTTGG ATAATAAAGA	7860
	TATTGCAACA CCATCAAGAA CGCGAGCGTT GTTAGATAAA TATGGCTTTA ATTTTAAAAA	7920
5	AAGTTTAGGA CAGAACTTTT TGATAGATGT GAATATCATT AATAATATCA TTGATGCAAG	7980
	TGATATTGAT GCACAACTG GGGTGATTGA AATTGGTCCA GGCATGGGGT CATTGACAGA	8040
10	ACAATTGGCC AGACATGCTA AAAGAGTATT GGCATTTGAA ATTGATCAAC GTTTAATACC	8100
	TGTATTAAAT GATACACTAT CACCTTATGA TAATGTGACG GTGATTAATG AAGATATTTT	8160
	AAAAGCGAAT ATTAAAGAAG CTGTTGAAAA TCATTIACAA GATTGTGAAA AAATAATGGT	8220
15	TGTTGCAAAC CTGCCGTACT ATATTACGAC GCCAATTTTA TTAAATTTGA TGCAACAAGA	8280
	TATACCAATT GATGGCTACG TGGTGATGAT GCAAAAAGAA GTGGGCGAAC GCTTAAATGC	8340
	TGAAGTAGGT TCAAAAGCAT ATGGTTCGTT ATCAATTGTC GTACAATACT ATACAGAGAC	8400
20	TAGTAAAGTA TTAACGGTAC CTAAATCTGT ATTTATGCCA CCACCTAATG TTGATTCAAT	8460
	AGTTGTAAAA CTGATGCAGA GAACTGAACC GTTAGTAACA GTAGATAACG AGGAAGCATT	8520
	CTTTAAGTTA GCAAAAGCAG CATTTGCACA AAGAAGAAAG ACAATTAACA ATAACATATCA	8580
25	AAATTATTTT AAAGATGGTA AACCAACAAA AGAAGTGATT TTACAATGGT TGGAACAAGC	8640
	AGGTATTGAT CCAAGACGTC GCGGTGAAAC GCTATCTATT CAAGATTTTG CTAAATTGTA	8700
30	TGAAGAAAAG AAAAAATTCC CTCAATTAGA AAATTAAATG ATTGACAAAG CAAAGCACTA	8760
	TTGTTAAAT TTAAATTTTG TTTGACGAAA ACGTTGCAA TATGGTATTA TGTAACTTGT	8820
	AGCGAGGTGG AGCAATATGC CAAAATCAAT TTTGGACATC AAAAAATTCTA TTGATTGTCA	8880
35	TGTAGGAAAT CGTATTGTAC TGAAaGCCAA TGGAGGCCGT AAGaaAACAA TAAACGTTT	8940
	TGGAATTTTA AAAGAAACAT ATCCGTCAGT TTTCATTGTT GAGTTAGATC AAGACAAACA	9000
	CAACTTTGAG AGAGTATCTT ATACATACAC TGATGTGTTA ACTGaAAATG TTCAAGTTTC	9060
40	ATTTGAAGAG GATAATCATC ACGAATCAAT TGCACACTAA ATAAGACATA TAGAGATGTT	9120
	AGACGTTTCT TAGTATAAGA AGTAAATATT ATGATAATTA TTTGAGTGTT GGGcATTATG	9180
	TTCAATACTC TTTTATTTA CAAAATGTTT AACACTGATG TTTCGCTTAT AGATTTTTCA	9240
45	GTAAATGGAT AATTGTATTT ATAAACACAA ATACAAGTAA ATACTAAGTA ATTAGATGGA	9300
	GAAAATTACT TTTTATTAA AAAAACACTA AAAAACAAAT TAAAATGTCA AATATTAATT	9360
50	CTCTTTATGT TAAAATCATC ATATTAAGAT AACGAAAAGA GGGCGGAAAA TGATATATGA	9420
	AACGGCACCA GCCAAAATTA ATTTTACGCT CGATACACTT TTTAAAAGAA ATGATGGCTA	9480
	TCATGAGATT GAAATGATAA TGACAACAGT TGATTTAAAT GATCGTTTAA CTTTTCATAA	9540

55

5 AAATCTCGCA TATCGTGCAG CGCAACTATT TATTGAGCAA TATCAACTAA AGCAAGGTGT 9660
 AACAAATTCT ATCGATAAAG AAATACCTGT TTCTGCTGGC TTAGCTGGAG GTTCGGCTGA 9720
 TGCAGCAGCA ACGTTAAGAG GATTGAATCG ACTTTTTGAT ATAGGGGCGA GTTTGGAAGA 9780
 ATTGGCTCTA CTAGGCAGTA AAATCGGGAC AGATATTCCG TTTTGTATTT ATAATAAAAC 9840
 10 TGCACATATGT ACTGGAAGAG GAGAGAAAAT CGAGTTTTTA AATAAACCAC CTTCAGCTTG 9900
 GGTGATTCTT GCTAAACCAA ACTTAGGCAT ATCATCACCA GATATATTTA AGTTGATTAA 9960
 TTTAGATAAG CGTTACGACG TACATACGAA AATGTGTTAT GAGGCCTTAG AAAATCGAGA 10020
 15 TTATCAACAA TTATGTCAAA GTTTGTCTAA TCGATTAGAG CCAATTCTG TTTCAAAACA 10080
 CCCACAAATC GATAAATTAA AAAATAATAT GTTGAAAAGT GGTGCAGATG GTGCGTTAAT 10140
 GAGTGAAGC GGACCTACTG TGTATGGGCT AGCACGAAAA GAAAGCCAAG CAAAAAATAT 10200
 20 TTATAATGCA GTTAACGGTT GTTGTAAATGA AGTGACTTGA GTTAGACTAT TAGGATAGAA 10260
 GGGTTGAAAA GATGAGATAT AAACGAAGCG AGAGAATTGT TTTTATGACG CAATATTTGA 10320
 TG 10322

(2) INFORMATION FOR SEQ ID NO: 99:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5614 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

35 GATTGATTAA ATGTTTTAAT CCACTTCAAT GCCTTCGATA AACTCTACAA TCGCGCTATT 60
 CATATAATTA TTCGATTTC A TTTGTTTCAGC ATATGTCTCA TTAAATCCAG ACATAACTTT 120
 40 TTAAAWGCG AAAATTGAAA TTGGTATCGT TACTAATAAG GCACTAGCCA TACGCCAATC 180
 AATGAGCATT ATGTATAAAA AGATAGCAGC TGACAAAAGT AAGTTTCCTA TAACTTCAGG 240
 AATCATATGT GCTAAAGGTA ATTCTATTGT TTCAACCTTA TCGACAAATA TATTTTTTAA 300
 45 TTCACCTATT TTCTTAGATT CCACTACGCC TAAAGGGAGA CGCATTAAAT TTTGAGCTAA 360
 TTTTTTACGA ATTTTCAGATA AAATTTTATA TGCCGTAATA TGTGATAGCA TCGTTGACGC 420
 TCCAAAACAA CAACTTGTG AAATATAAGC GATTAAAGCA ATAAAGATAT AAACCATAAT 480
 50 CGAATTAATC GTATATGTAT TGTTAATCAT CATTAAATA ATTTTAAATA CTGCCCAATA 540
 AGGAACTAAT CCAGAAAAGA CACTGATGAT AGACAACAAA ATTGATAACA TAATTTTCCA 600

	ATATGTAACCT CCTkTCAATT AATAATCTAA ATTAAGCCGC TTATATTATT TATTTCACTG	720
	GATGATATAC ATAATATAAA TTTGTTATTT GTTAAAAATT AATACTTATT ACAAGTACAT	780
5	CATATATTAG TTGATAACGA TTATCAATGT CGCGTGGATT TGTGACACAT TTCTTTTAAA	840
	AATTCACAAG GTTATGGGGC AGAAATGATA AAGAGCCACT AATGATTTAT TATGTAGTGG	900
10	TTCTGGGAGT GGGACAGAAA TGATATTTTC ACAAATTTA TTTTCGTCGTC CCACCCCAAC	960
	TTGCATTGTC TCTAGAAATT GGGAATCCAA TTTCTCTTTG TTGGGTCCCT GAATATAGCC	1020
	TTGTAGAGTC TAGTACATTG ATTTGTATCC CAATGTCCCT ATAATTGATT ATTCGCTTTA	1080
15	TCTAATGATC CTATGACTCA ACTATTAAAT CATTTTTCGA AATACTTAAT TCTAATATAA	1140
	TTAAATTTCAT TTATTGTAAT ATTGCAAAAA TACATTGCAC ACCTTGTTCA TCAATGCTAT	1200
	AATTAATTAC ATAATAAATT GAACATCTAA ATACACCAAA TCCCCTCACT ACTGCCATAG	1260
20	TGAGGGGATT TATTTAGGTG TTGGTTATTT GTCACCTTTT TTATTGTTGC GCGTTCGTAA	1320
	CCAATGTGCA AAAAACGCAA CAAGACAGCC GCTTATAGCT GAAGTCATGA TGTTAATTAA	1380
	TAAATTGAAC ATCCGTCATA CACCTCCTCT CTGCGTTAAA GTAACGCCCG AGATGTTAGG	1440
25	CGACCATCAT ATTATATCAT TTATTTATTA TATTTACGC AATATTAAGG CTTAAGTAAA	1500
	GTTTTTTTTTA GTGGTTTACG CTACTTTAAT TGCTATCTTT TAAAATCCAT TTAGATAATA	1560
30	TAAATGTGAT GGGTATCGTA ATAATTAAAC CAGCAAATGG TGCAATTTCT GCTGGCAAAT	1620
	TTAGCCAGGA TACAAATACA TATAATAAAA CTGTTTGTA GCTTACGTTG ACAATCTGCG	1680
	TAATTGGAAA ACTAATGAAT TTTCTCCAAG TAGGTTTTAC CCTGTAAACA AAATAACAAT	1740
35	TCAAATAATA TGAAATCACA AAAGCGACTA GAAATCCGGT AATATGACTA ATCATATATT	1800
	CAATGTGTAA TAATTTTAAC AGCAATAAAT AGACAACATA ATAATTTAAC GTATTAATGC	1860
	CGCCAACAAT GATAAATTTT AAAATTTTCAG CATGCGTTTG TGTTAGTTTC ATATGTGTAC	1920
40	TCCTCAACAT CAAAATATAT GCATAACTAC GTTCTCGAAC ATACTCGAAT ATGCGAGCCA	1980
	ATCCGCTTCA CTTCAAATAT GCTTATTTCA ATCTTTATAC CCTTTCACAG CAAATTTAGT	2040
45	CTCTTTCCCC TCATCCTTAT ACGCCATTAT AATGTAAGT ATTTATCGCG TGA CTCATTA	2100
	GCACTATAGA GATTACTTTA GTTCACTAGT AATTTTATAT ACAATAAGAG CGACAACAGT	2160
	AATGAGAGGA TGTCTACTAT GCAATTACAA AAAATTGTCA TCGCTCCTGA CTCATTTAAG	2220
50	GAAAGTATGA CCGCACAGCA AGTTGGCAAT ATTATAAAAC AGGCTTTTAC TAATGTTTAT	2280
	GGGAATACCC TTCATTATGA TATCATTCGG ATGGCTGATG GTGGTGAAGG TACCACAGAT	2340
	GCTTTAATGC ATGCAACAGG TGCCACTAAG TATACAGTCA TCGTTAATGA CCCTTTAATG	2400

55

	GCGGCAGCGT CAGGTTTGA TTTATTAGAA AAAGAGGAAC GTAATCCTTT ATACACATCA	2520
	TCATATGGTA CCGGTGAACT AATTAAAGAT GCATTAAATC ATGGTGCTAA GACCATTATT	2580
5	TTAGGGATTG GTGGCAGTGC AACAAATGAT GGTGGTACAG GTATGCTAAG TGCACTAGGC	2640
	GTAAAGTTTA CTGATGTAAA CGGGGACTTA TTACAAATGA ATGGTGCTAA TCTTGCTCAC	2700
	ATTGCACAAA TCGATATAAC CAATCTAGAT TCGCGATTAA AAGAGGTGAC CTTTAAAGTG	2760
10	GCCTGTGATG TTTCAAATCC TTTATTGGGT GAAAATGGTG CTACCTATAT TTATGGTCCT	2820
	CAAAAAGGCG CTGATGCAAA GATGATACCA AAGTTGGATT TCGCAATGTC GCATTATCAT	2880
	GATAAGATAA AAATGTGCAC AGGAAAGTCC GTTAATCAAA TACCAGGTTT TGGTGCAGCT	2940
15	GGCGGTATGG GCGCAGCATT ATTAGCGTTT TGTGAGACAA CTTTAACAAA AGGTATTGAT	3000
	GTCGTCTTTG ACATTACAGA TTTTCATCAA AGAATTAAAG ATGCAGACCT CGTTATTACT	3060
20	GGAGAAGGAC GCATGGATTA TCAGACCATC TTTGGTAAAA CACCCGTAGG CGTTGCGTTA	3120
	GCTGCAAAAC AATATCATAT TCCTGTCATC GCGATTGTG GCAGTCTAGG CGAAAATTAT	3180
	CAACATGTTT ACGATTTCGG TATTGATAGT GCCTATTCTA TAATCTCTTC ACCTAGCACT	3240
25	TTAGAAGATG TCCTACAAAA TAGCGAACAA AATTTATTAA ACACTGCAAC TGACATTGCT	3300
	CGTATTCTGA AATTACAATA ATGTCAAAGT AAATCATCAG CTTTATTATT TGCAGTTAAA	3360
	ACTTGAATGA GGTGAAACCC ATGAAAAGAA CTGATAAATA CCGTGATTCA TATCAATACG	3420
30	ACAATCAAAA CCAAATCAT CGTCGTCAAT CTGAAGACGC ATCGTATAGA CAACAATATG	3480
	CTAAAGGCGA TCCTGAAGAA CACCCGGAAC GATACTATAA TGGTAGAGAT TATCGAAGAG	3540
	AACAAATTCT TGAAGAAGAA AACGAGAAAT CCCGCCGTTT AAAAAAATGG TTATATATCA	3600
35	TTATTGCCAT TCTCTTAATT ATTGTCGCTA TTTTGTGCAC ACGCGCCTTA CTTAACAATG	3660
	ATAGCGATAA AGTTAGTAAT GACCCTAAAG TCTCTCAAAA TTATAAAAA CAAGTTGAAA	3720
40	ATCAAGACGG CCAAATTAAC CAGCAAGTAG ATAATGCTAA AGAAAATATT AAAACAACC	3780
	AAAAAACTGA TGACATTATT AAAAATTTAC AAAATCAAAT CGACAACCTG AAGCAGCAAG	3840
	AACAAAACAA AGCTGATTCT AAGCTAACTC AATTTTATCA AGACCAAATC AACAAATTGA	3900
45	CAGAGGCAAA TAATGCACTT AAAACAATG CAAGCCAAGG TAAAATTGAA AGCATGTTAA	3960
	ATGATATTAA TACAAAATTC GACAGTATTA AATCTAAATT AGAAAGCTTA TTAAAGATG	4020
	ACAATGGTGG CGCTAATTAA TTATTACACC TGCTTTGATG ATAAACATTA ATTCCCTATA	4080
50	CTTTATCTGT ATCACTACGT TATTCGTGAT GATGCATTAA GAGTATAGGG ATTTTTTATA	4140
	TAACTTGTA TTCTAACTAC ATACAAATAC ACACAAAACG TATATAATTT ATATAATTAT	4200

55

	TTATTGCTAA TTACGTTAGG CGTCATGACC GCTTTTGGCC CACTAACTAT AGATATGTAC	4320
	GTACCATCAT TACCTAAAGT GCAAGGTGAT TTTGGTTCTA CTACATCAGA AATTCAATTA	4380
5	ACATTATCAT TCACAATGAT TGGTCTTGCA CTAGGCCAAT TTATCTTTGG ACCTTTATCC	4440
	GATGCTTTTG GTCGCAAACG GATTGCTGTA TCCATTTTGA TCATTTTCAT TTTGGTATCA	4500
	GGTTTGCTCA TGTGTGTTGA TCAATTGCCA TTATCTTAA CTTTACGATT TATTCAAGGT	4560
10	TTAACTGGTG GTGGCGTCAT CGTGATTGCA AAAGCCTCTG CTGGTGATAA ATTTAGTGGC	4620
	AACGCACTCG CTAAATTTTT AGCATCTTTA ATGGTAGTTA ATGGCATCAT CACTATTCTT	4680
15	GCACCATTAG CCGGTGGATT AGCTTTATCC GTAGCAACAT GGCGTTCTAT TTTCACAATT	4740
	TTAACTATTG TGGCACTCAT CATTTTAATT GGCGTCGCTT CTCAATTACC TAAAACATCT	4800
	AAAGATGAAT TAAAGCAGGT GAATTTTAGT AGCGTCATTA AAGATTTTGG AAGTCTTTTG	4860
20	AAAAAACCAG CATTTATTAT TCCAATGCTA TTACAAGGT TAACTTATGT AATGCTATTT	4920
	AGTTATTCAT CTGCATCGCC ATTTATTACT CAAAAATTGT ATAATATGAC ACCCCAACAA	4980
	TTTAGTATCA TGTGTGCTGT TAACGGTGTA GGTTTAATCA TTGTCAGTCA AGTCGTTGCT	5040
25	TTATTAGTAG AAAAATTACA TCGCCACATA TTATTAATCA TTTTAACTAT TATACAAGTG	5100
	GTAGGTGTTG CTTTAATTAT CCTGACACTT ACATTCCATT TACCACTTTG GGTCTTACTC	5160
	ATCGCATTCT TCTTAAATGT GTGTCCTGTG ACGTCAATTG GACCGCTTGG TTTCACAATG	5220
30	GCTATGGAAG AACGAACAGG TGGCAGTGGT AACGCATCAA GTTTACTTGG CTTATTCCAA	5280
	TTTATCTTAG GTGGCGCTGT TGCACCATTA GTTGGCTTAA AAGGCGAATT TAATACATCA	5340
35	CCATATATGA TTATTATCTT CATTACAGCC ATTCTATTAG TCAGTCTACA AATCATTTAC	5400
	TTTAAATGA TTA AAAAGCA ACATGTCGCA TAACACTTCA ACATAATTAG AACCTAGCA	5460
	AAGATATCTA TCTTTGTCAG GGTCTTCTT TATGAATTAT GAGATCGAAT CTTCAACTAA	5520
40	AATTACGCCT TCATAGCAAG GACATTTCTA TTCAATCACC CTTTAACAGG CATCCAAATT	5580
	TcTGTAATAT ATTTTTCAC TGTAGTATCA CCAT	5614

(2) INFORMATION FOR SEQ ID NO: 100:

45	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 9179 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

55

	AAAGACAATG ATATGAAGTA TATGGATATC ACAGAA ^a AAAG TGCCAATGTC GGAATCTGAA	120
	GTAAACCAAT TGCTAAAAGG TAAGGGGATT TTAGAAAATC GAGGGAAAGT TTTTCTAGAA	180
5	GCTCAAGAAA AATATGAGGT TAATGTCATT TATCTTGTTA GCCATGCATT AGTAGAAACA	240
	GGTAACGGCA AATCAGAATT AGCAAAAGGC ATTAAAGATG GGAAAAACG CTATTACAAC	300
	TTTTTTGGTA TAGGAGCATT CGATAGTAGT GCTGTTTCGT GTGGGAAAAG TTATGCTGAA	360
10	AAGGAACAAT GGACATCACC AGATAAGGCG ATTATTGGTG GTGCAAAGTT CATTGCTAAT	420
	GAATATTTTG AAAACAATCA ACTGAATTTA TATCAAATGC GATGGAATCC AGAAAATCCT	480
	GCGCAACATC AATATGCGAG TGACATTTCG TGGGCAGATA AAATTGCCAA ATTAATGGAT	540
15	AAATCCTATA AGCAGTTTGG TATAAAGAAA GATGATATTA GACAAACATA TTATAAATAA	600
	GACATCGGTG CTTAAAGGAG CTGGAACAAT TTATGTTTC GAGCTCCTTT AGCGCATTCT	660
20	GAGTGTGTTA GTTAAATGGA TTTTAACCTA ACAAAAAACG CTATATAGCA TCAAATATGC	720
	TATATCCAC ATCATTGTTA CAAATGTACA TGATGTAAAT GAATATTGCT GTCTAAATGT	780
	GCATGTAATA TACAATGGTG CAGATAATAC ACTTAAAGTCC TTAAAAATGA AACGTTAgTT	840
25	CCAAGAGTCA TTTT ⁻ TAAACA ATAGTGCATG TGATAAAATA GAAAAGAATG AAAAATATAG	900
	AGGTGACAAT ATGAAGATAG CAATTATAGG TGCAGGCATC GGTGGATTAA CAGCTGCTGC	960
	ATTATTACAA GAACAAGGTC ATACTATTAA AGTCTTTGAA AAAAATGAGT CAGTTAAAGA	1020
30	AATTGGCGCT GGGATTGGTA TCGGAGATAA TGTGCTTAAA AACTAGGTA ATCATGACTT	1080
	AGCTAAAGGT ATTA ⁻ AAATG CTGGGCAAAT CTTATCTACA ATGACAGTGT TAGATGACAA	1140
	AGATCGCCTG TTA ⁻ ACTACTG TTAAATTAAA AAGTAATACA TTGAATGTGA CGTTACCACG	1200
35	CCAAACATTA ATTGACATTA TTAAATCTTA TGTA ⁻ AAAGAT GACGCAATAT TTACAAATCA	1260
	TGAAGTCACG CATATAGATA ATGAGACAGA TAAAGTTACC ATACATTT ⁻ CG CGGAACAAGA	1320
40	AAGTGAAGCA TTTGATTTAT GTATTGGTGC TGATGGAATT CATTCTAAAG TGAGACAATC	1380
	TGTAAATGCT GACAGTAAAG TATTATATCA AGGGTATACA TGCTTTAGAG GTTTAATTGA	1440
	TGATATTGAT TTAAAGCATC CGGa ⁻ TTGTGC AAAAGAATAC TGGGGa ⁻ AGAA AAGGa ⁻ AGAGT	1500
45	AGGTATTGTT CCGTTATTAA ATAATCAAGC ATATTGGTTC ATTACAATTA ACTCGAAGGA	1560
	AAACAATCAT AAATATAGTT CGTTTGGTAA ACCTCATTG CAAGCATACT TTAATCACTA	1620
	TCCAAATGAA GTTAGAGAGA TCTTAGACAA ACAAAGTGAA ACAGGTATCT TATTGCATAA	1680
50	TATTTATGAT TTGAAACCAC TCAAATCTTT TGTTTATGGT CGTACTATTT TACTAGGAGA	1740
	TGCAGCACAT GCGACAACGC CTAATATGGG GCAAGGTGCT GGACAAGCAA TGGAAGATGC	1800

55

	TAAAATACGT GTCAAACATA CTGCAAAAGT AATTAAGCGT TCTAGAAAAA TCGGTAAAAAT	1920
	TGCCCAATAT CGTAGTCGTT TATTGTGTC AGTTAGAAAT CGTATTATGA AAATGATGCC	1980
5	AAATGCATTA GCAGCTGGAC AAATAAATT CTTATATAAA TCGAAAGAAA AATAATACAA	2040
	CAATATGAAA ACCCCCGTAT GTTGAAACGA GAGCTCAACA TATGGGGGTT CTTGTTTTTA	2100
10	TAATGTTATT ATAATAAATT CAATTATTAG TTAACGACAA ATTGTGGTTT CTCACCTTGA	2160
	ACGGCACTAA TTGCAGCATT AGCAACAATT TTAGACATCA TGTCACGTGC TTCAAATGTA	2220
	GCATTACCAA TATGCGGTGT TAATACTACA TTATTAAGTG ATTTTAAGTC ATCGGTAATA	2280
15	TCTGGTTCAA ATTCATATAC ATCAAGTGCA GCACCTTCAA TTTCATTATC TTCAATGCT	2340
	TGCACTAGTG CTTGTTCGTG CACGATTGGA CCACGAGAGG CATTGATTAA ATACGCCGTA	2400
	GATTCATCA TTTTAAATTG TTCTGTATCA ATTAAATGAT GCATTTTAGG ATTATAAGCA	2460
20	GCGTTGATAG TGATAAAATC TGCATTCTTT AATAGTGTAT CTAAATCTAC ATATTTTGCA	2520
	CCGATTTCTC GTTCTTTTTT TTCTTTGCGA TTAGGTCCAG TGTATAGCAC ATCCATGTCA	2580
	AATGCTCTTG CACGACGAGC TACTGCACTA CCAATTTTAC CTAAACCGAT AATGCCGATT	2640
25	GTTTTCCCAG ATACTTCTCT ACCTCTGAAA AATAAAGGTG CCCATCCATC AAATCCAGTT	2700
	GTACGTGATA ATTGGTCCCC TTCAACAATA CGACGCGCTA CTGCAAGTAC TAATCCAATT	2760
30	GTTAAATCAG CAGTCGCGTT TGTGATGCT TTAGGTGTGT TTGTAACATC TATACTTTTT	2820
	TCTCGGGCAT ACTCGATATC AATATTATTA AAACCAGCGC CATAGTTGGC AATGATTTTT	2880
	AAGTCTTTAC CAGCATCGAT AACATCTTTA TCAACGTTTG TAGATAATAA ACTAATTAAG	2940
35	GCAGTCGCGT TTTTAACACC TTTAATTAAA GTGTCTTTAT CGACTAATCC TTTACCTTCA	3000
	TACATTTCAA CTTCAAAATG TTCTTGTAAG AGTTTTAAAC CTACTTCTGG TATtGCACCA	3060
	gCAACATAAm CTMTtCCAT AAAAGAtCAC TCCTTTTATC TTAGTATAGT AGAAGATTAG	3120
40	ACAGTATACA ACTATGTCAT GATGTCTTGT GTATCAATGA TGTAAGCGCG TACTTTTGAT	3180
	GGAGGCGATA TAACTTAGGC ACTGTAGAAC TATGAATATT GTAATGTGGA AAAACTGGAT	3240
	CAATTAAATT AGATAACGTA GTTTTAAAGT TAATAGTATT AGAAAAAATT AATATTTTGA	3300
45	ATATGGGAGG AAATATAAAT AAGTAGGTGG CAACGAAAAA TAGCAAAAAA AGAGCTTCTC	3360
	CTATAAAGGA AAGCTCAAAG TTTTTTGATG ACATATGTAC TAGAATTAAG TTTCAAGACA	3420
	ATATGTATCA TCGTGTtTAT ATTAAATATG GATGTAGTTG TAGTTACCTG CTTCACTTGC	3480
50	AGAAATAGTT CTAGAACTTA CTGAGAAAGG TCCGCCACTA TAATTCATTT CTGAAATTGT	3540
	AACTGAACCA TCACTGTTTA CACTTTCTAC ATATGCAACG TGACCAAATG GTCCTTCAGA	3600

55

	AGCAGCAGCC CAATTATTAG CATTTCCTCCCA AGTAGAACCG ATTTCTCCGC CAACTTTATC	3720
	ATATACATAC CAAGTACATT GTCCTGCAGT GTATAAGTTA CCAGAATGTG AAATTGATGA	3780
5	TGTAGTTGTC GTAGTTGTCG TAGTCGTTGT AGTTTGAGTC GTGTTGTAGT TATAGTTGTT	3840
	GTAATTTGTA TAATTTTCAG CAGCATCTGC ATGATGTGCT TGACCTACTA ATGCTGTGCC	3900
	GATTCCTGCT GTTAACGTAG TTGCTGTTAC TAATTTTTTC ATGAATAAAG TCCTCCAAAG	3960
10	TTCTATATCT TTTTTTATAA ATAAAACGTA GCGACTGTTT TATTCTCACA TCTCGAATTG	4020
	ATGACAATAG TTACTTTAAC AAAATTAATG CTTCTTGTGG GGAATGTTAT TGATTTGTAA	4080
15	AAGAATAAAA AAACTTTGAC TAATTTTGTA ATAAAAATTA GTCAAAGTTA CAATGAGATT	4140
	AACAGATAAT TAATAGGAAA TATTTATTTG TAATATGTTT AAATAAATCG AATTGTTAAA	4200
	GGTATTATAT ATTCTTGGCC ATTATAATAT TTGACACACG CAATAATTGT GAATACAAAA	4260
20	GATAATATTG AGAAAGCGAA TATGGATAAA ATACCGATAA ACGTAATGAT GAAACCTATA	4320
	ATAATAATGA AATCAATATC TGTAGCAATT AGGAAAACGC CTATTAAAGT GATAACGACT	4380
	AAAACGATAG ACCAAATAAT ATAAGAAATC GTATAGTTAA GATAATTTTT TCCAGCACGA	4440
25	TCAACTAGTT TCGATTCATC TTTTTTCAAT AACCATATTA TCAGTGGACC AATAATAGAT	4500
	GTGAATAAAC TTAATAAATA GATAAGCATC GCCATAATGT TCTCATCATT GGATTGCGA	4560
	TTCGGTTGAT GATTTGTTAC GTCGTTCAAT TCAGTTGTCA TATTAGACAC TCCTTTGAAA	4620
30	ATTGTAATAT TATCTTTAAC TATAACAAAA TATAATCAAA AATAAACATG TTTATTAAAC	4680
	AATTATTAAA AATAAAAATA ATTGGTGGAC GTCGCGGTTT AAATAGGTTA ATTTAAGGTT	4740
	ATATATACTT AACATTTATA ATGATGCGTA ATGAATTCGC ATCATTTTTA TATTGTCTTA	4800
35	CGTATAATTT GTTTTTAATT TTAACCAAAG ATAGAAAGAG GGTTGTTTAT GAAAATAGCA	4860
	ATTGTAGGAT CAGGAAATGG CGCAGTTACG GCAGCAGTAG ATATGGTGAG CAAAGGCCAC	4920
40	GATGTTAAAT TATATTGTCG TAATCAATCT ATAAGTAAGT TTCAAACGC AATCGAAAAG	4980
	GGCGGATTTG ATTTTAATAA TGAAGGTGAT GAACGTTTCG TAAAATTCAC TGATATTAGT	5040
	GATGATATGG AATATGTTTT AAAAGATGCT GAAATTGTTT AAGTGATTAT TCCATCTTCA	5100
45	TACATAGAGT ATTATGCTGA TGTAAATGGCA GAGCATGTAA CTGATAATCA GTTGATATTC	5160
	TTCAACATGG CTGCAGCAAT GGGGTCAATT CGTTTTATGA ATGTTTTAGA AGATAGACAT	5220
	ATTGAAACAA AACCACAAC AGCGGAAGCT AATACGTTGA CGTATGGTAC GCGTGTGAT	5280
50	TTTGAAAATG CAGCAGTTGA TTTATCTCTA AATGTACGTC GTATCTTCTT TTCAACATAT	5340
	GATAGAAGCT GTCTAAATGA TTGTTATGAC AAAGTTTCAA GTATTTATGA TCATTTAGTA	5400

55

	CCAACATTAT TGAATGTCGG TCGCATTGAT TATGCTGGCG AGTTCGCTTT ATATAAAGAA	5520
	GGAATTACTA AACATACAGT TAGATTACTT CATGCAATCG AATTAGAACG TTTGAATTTA	5580
5	GGCCGTAGAT TAGGTTTTGA ATTATCAACA GCTAAAGAAT CACGTATTGA ACGTGGTTAT	5640
	TTAGAACGTG ATAAAGAAGA TGAACCATT AATCGTTTGT TTAATACAAG CCCAGTATTT	5700
	TCACAAATTC CAGGACCAAA TCATGTAGAA AGCAGATATT TAACTGAAGA TATTGCATAT	5760
10	GGTTTAGTAC TATGGTCAAG CTTAGGTCGT GTTATTGATG TACCGACACC AAATATAGAT	5820
	GCAGTAATTG TAATTGCATC AACCATTTTA GAGAGAGACT TCTTTGAGGA AGGCTTAACA	5880
15	GTTGAAGAAA TTGGTTTAGA TAAGCTTGAT TTAGAAAAAT ATTTAAAATA AATGATGGCT	5940
	TGAAGATAGA AAAGGATATA GCATTATGCA AAAGCAATAA ATTGAAGAAA AGAGGTTTCT	6000
	CATCAATAAG CGnAGGGGAC GATAGATGAT GAAAAGAAAA CCCACCTTTT TAGAATCAAT	6060
20	TTGACAATG ATTGTAATGG TTATTGTTGT TGTAACAGGC TTTGTGTTTT TTGATATTCC	6120
	AATTCAAGTA TTATTAATTA TTGCCTCAGC ATATGCCACA TGGATTGCAA AACGTGTAGG	6180
	CTTAACATGG CAAGATTTAG AAAAAGGCAT TGCAGAACGT TTAAATACTG CAATGCCTGC	6240
25	AATTTTAATT ATACTAGCGG TAGGAATTAT AGTAGGCAGT TGGATGTTTT CTGGCACAGT	6300
	GCCAGCCTTG ATTTATTATG GCTTAGATTT ATTGAATCCA AGCTATTTTT TAATATCAGC	6360
	CTTTTTTATA AGTGCTGTTA CATCTGTAGC AACTGGTACA GCATGGGGCT CTGCATCAAC	6420
30	TGCAGGGATT GCACTTATTT CTATTGGTAA TCAATTGGGG ATTCCTCCAG GGATGGCAGC	6480
	GGGTGCTATT ATAGCAGGGG CTGTGTTTGG CGATAAAATG TCACCATTAT CAGATACAAC	6540
35	TAATTTAGCG GCGCTTGTTA CTAAAGTTAA TATATTTAAA CATATACATT CGATGATGTG	6600
	GACGACGATA CCTGCATCAA TCATAGGTTT ATTAGTATGG TTTATTGCTG GATTTCAATT	6660
	TAAAGGGCAT TCAAATGATA AACAGATTCA AACTTTGTTA TCAGAGCTTG CACAGATTTA	6720
40	TCAAATTAAC ATATGGGTCT GGGTTCCCTT AATTGTGATC ATTGTTTGTT TGCTATTTAA	6780
	AATGGCTACA GTGCCAGCTA TGCTAATATC AAGCTTTTCT GCCATTATAG TGGGGACTTT	6840
	TAATCATCAT TTCAAATGA CAGATGGTTT CAAAGCAACA TTTAGTGGTT TTAACGAATC	6900
45	AATGATACAT CAGTCTCATA TTTCATCCAG TGTGAAAAGC TTGTTAGAAC AGGGTGGTAT	6960
	GATGAGTATG ACCCAAATAT TAGTAACGAT ATTTTGCGGA TATGCATTTG CAGGTATTGT	7020
	AGAAAAAGCA GGATGTTTAG AAGTCTTATT AACTACTATT TCTAAAGGCA TCCATTCTGT	7080
50	AGGAAGTTTA ATATGTATTA CTGTTATTTG TTGTATTGCG CTTGTATTCTG CTGCAGGTGT	7140
	TGCTTCGATT GTAATTATTA TGGTCGGTGT GTTAATGAAA GATTGTTCG AAAAAATACCA	7200

55

	AATACCATGG GGAACATCAG GTATTTACTA TACGAATCAA CTTCATGTCT CTGTTGAAGA	7320
	ATTTTTCATA TGGACAGTAC CATGTTATTT ATGCGCAATT ATAGCAATTA TCTATGGTTT	7380
5	TACAGGGATA GGTATTAAAA AGTCATCGAA TTCACGTTTA ACTTAATGTG AGCGTGGAAT	7440
	ATATATAATA TGTTGAAACA CTTTAATCAT TTATAATTGT AGCGGTTATA ATTTGAAAAG	7500
	GTTTTAACTT AGAATAAATA TCCTCTATGC ATATACTGAA TATGTTTTGT AGCGGAACAT	7560
10	GTTGATATAT GTAATGTAAG TTTTATGTCA TGATTGTAA TGAATAAATT AATTGAGAAT	7620
	TTGAAGGCAA GTATATTTGT AAGTACTTTA ACTAAAAATT TATCAATGTA TAGCCGATTT	7680
	GACATGCCTA AATTTGGGTG TGTCAATGGC TGTATGTTGT TTATTCTTTA TTACAGAGTG	7740
15	AATCGGATTG GTGAAAATCG AAATTTTGAG ATTTTACCA ATTCGATTTT TTTCATAGAA	7800
	ATTAATAAAG CCAACAAGGC TCTTGAAACC TTGTTGGCGT AAACATAGCC ATCACTAATT	7860
20	AGTGAATGAA GTTATAACCA GCAGCTTGGC TAGCTGAGAT TGTACGTGAA GTTACAACAC	7920
	CTGGGCCATA ACCATAGTTC ATTTCTGAAA CTCTTACTGA ACCATTGCTG TTAACACTTT	7980
	CAACGTATGC AACGTGACCG TATGCACCTT GAGTTGTTTG CATAATTGCA CCAGCTTTTG	8040
25	GTGTATTGTT CACTGTGTAA CCAGCTCTTG CAGCTGCGTT AGCCCAGTTA CTTGCATTGC	8100
	CCCAAGTTGA ACCGATTTTA CCACCTACAC GATCAAATAC GTAGTATGTA CATTGACCAG	8160
	AAGTGATAA GTTACGTCCT GAAGTATAAC CACTTGAGAT TGAACGGCCA TTTGATGATG	8220
30	GAGCCATAGT TGTAGTTACT TGAACATTGT TGCTTGAAGT GCTGTAGCTT GCACCTAAAC	8280
	CACCACTAGC GTAGCTGTTT GTGTTGTAAC TATTATAGTT ATTGTAGTTA TATGATTGAT	8340
	TATTATTTGA GTAGTTGTTG TAACGGCTGT AGTTATTGTA GCTATAACCG TTGTTGTAAT	8400
35	TGTTATAGTT ATTGTAACCA TTGTAGTAGT AATAGCTGTA GTAGCCATTA TCTTGTTTTA	8460
	ATTGACTTGG ATGCCAGTTA CCTTTCCATG TGTAATGGTA GTTACCTTGT GCATCAATAG	8520
	TGTAAGTATA GCTATATGAT GTTGGGTCGT TTGGATTATA ACCGTAGTTA TCTTGCTCAG	8580
40	AAGCATGAGC TTGATTTCTT GATGCAATTG CGATTGTAGC GAATCCTGCA GTTGCGATAG	8640
	TAGCTGTAGC GATTTTCTTC ATTTTAAAAA TATCCTCCTA AAAATTTTAA ATCTAAAATA	8700
45	TTTTCGTAAT GTCCGTGTGA CAAAATTAAT GTTATAAGTT ATCTCTCGTA ATTAAACGAC	8760
	AAGAAAGACT ATAACAGAAA TTAGCGTCCT TGTGTGCTTT GTTAACGTTT TGTAATTTTT	8820
	TGCTAATATC TTGACACAAT AGAATTTTAA AAGTATAGAA ATTTGCATTT TGCAAAACTT	8880
50	ATAACTACGG CATTCTTTGT GAAAACTGAA TGTTTCGAAA ATAAGTCTGT TACAAATTTG	8940
	TAATATTACT GAAAATTCTA AATGTATATT TTGTGCATAA TATAGGACTT TTAATCAGAA	9000

55

GGATGAAAAT GTATATTTAA TGGATAAAAT ATCCTAATTT AGCATAAAAA AATGTTTTTAA 9120
TAAAAGTATT ATTTGATATA ATCGATTAT GTTTTGTAC TGCTAAAAAA CATGTGGCG 9179

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1868 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

CCTTCAGCCA TTTGACTTCG ACATGAGTTG CCTGTACATA TAAAATAAAT TGTTTTTTTA 60
GTCATAACAA TCTCCTAATT AATTAAAATA TGATAAGTGT TAGATACAAC CCTATGAGGG 120
TTATAAATAG TACTGGAATT GTAATGATGA TACCAGTTT AAAGTATGTG CCCCAAGAAA 180
TCTTAACATC TTTTGTGTT AAGACGTGTA ACCACAGTAA TGTAGCTAAA GAGCCTATCG 240
GTGTAATTTT TGGACCTAAA TCAGAACCGA TAACATTCGC ATAAATTAGG CCTTCTTTTA 300
ACATGCCATG GACATTTGAT TGACCAATAG CAATCGCATC TATTAAACT GTAGGCATAT 360
TATTCAATTAT TGATGATAAA AACGCTGAAA TGAAGCCCAT TCCCAAATA GTGCTAAATA 420
GACCGTAATT GGAAATATAT TCTAATATTT TAGCCAATAT TAAAGTAATG CCAGCATTTT 480
TTAAGCCGAA TACGACGATA TACATACCAA TTGAAAATAA TACTATATTC CAAGGTGCGC 540
CCTTAATGAC TTGCTTAATA TTTACAGCAT TTGATTACG AGCCAACATT AGAAAAATAA 600
AAGCAATGAT TCCAGTGAAA ATTGATACCG GAATTTTAGT AAATTTACTG ATTAGATAGC 660
CGAAAAGTAA TATAACTAGA ACAATCCaTG AAATTTTAAA TAGCTTTAAA TCATTAATGG 720
CATCFTTAGG ATGCTTTATA TTATTATCAT CAAACGTTTT AGGTATCGCT TTTCTAAAAT 780
ATAACCACAA TACTATAATA CTTGCTAAAA GCGAGAATAA ATTAGGTATA ATCATTCTAC 840
TAAATATCG AACGAATCCT ACATGAAAAT AATCAGCAGA TATAATATTC ACTAGATTGC 900
TCACGATTAA AGGTAAAGAA GTTGTGTCAG CTATAAAACC ACTCGCAATA ATnAAAGGGA 960
ATATGGCCCG CTTACTAAAA CCTATATTTT TAACCATCGC TAATACAATA GGCGTTAAGA 1020
TTAAcGTGCG CCATCATTTG CGAAAAATGC AGCAACAATG GCACCCAATA ATATGATATA 1080
AACGAACATT TTAAACCAT TGCCTTTTGA AGCATGAAGC ATGTGAATAG CTGACCATTTC 1140
GAATAATCCA ACTTTATCTA ATATTAATGA AATAAGAATG ACTGAGACAA AAGTCAAAGT 1200
AGCATTCCAA ACAATACCTG TTACTTCGAA AACATCGGAA AACTTTACAA CACCAGTAAT 1260

TAATACAAAT AATAAAGTTA CTAGAAAAAT GAGTGTGCT AAAGTTGTCA TCATTAGCAT 1380
 TCACCAGTCT TAAGGTTATG ACAAATACAT CGTTGGTTAG AGGTATGAAC CTTAGACAAG 1440
 5 TTATTAATTA CGGACTCAAA AATATTATGA TTgAGCTGGT ATAAATGTTT ATTTCCGATT 1500
 TTTCGTGTCG TAACTAAGTT GGTTTTTACT AATGCTTTCA TATGrTAGCT AAGTGTAGGT 1560
 TGAGAGAATT GAAAATGTGC TAACAAATCA CAAGCGCATA ACTCTCCACA AGAAAGTAAA 1620
 10 TCTAGTATTT CTAATCTGCT TGAATCTGAT AAAACTTTTA AAAATGTTGC TAGTTCTTTA 1680
 TACGTCATAA CATACTCTCT AGACGTTAAA TAGATTATCA TCTATATAGA TGAATGTCTA 1740
 TGTTCCCTTG GTATATTACA CGATATGACT ATGTAATTTA AATTTGGTTT TAGTATTAAA 1800
 15 AGGGTATTAA AGATAAATTA TAGATATTGA TTTTGCAAAA TATACTCTTT GTTCTGCATT 1860
 GAAAAAGG 1868

20 (2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15249 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:
 30 ATTTATGAAA TCCATAGCnA TAAACATTAT TCTTGCATCG GCTATACAAA CAGTTACCGC 60
 AAGCAAATTT GTATATCAAC CTGGAATTGT GTTCACGTCA ATGGCaAATG CCGATGATGT 120
 GTTATCAGGC GATAGTTATT TTATGGCTGA ATTAAAATCT ATTAAGCGTA TTGTTGAAAT 180
 35 TCCAGATAAT CAAAAAATAT ACTGCTTTAT AGATGAAATT TTTAAAGGTA CCAACACAAC 240
 TGAACGAATT GCCGCTTCAG AATCAGTACT ATCATTTTTA CATGAAAAAT CTAACCTTTAG 300
 40 AGTTATTGCA GCAACACATG ATATTGAGTT AGCTGAACTC TTTAAACAAC GTTATGAAAA 360
 TTACCATTTT AATGAGGTAA TAGAAAATAA TAACATACAT TTTGATTACA AAATTAAGCC 420
 TGGCAAAGCA AATACACGTA ATGCCATCGA ATTATTAAAA ATCACTTCAT TTCCAGCAAA 480
 45 AATATATGAA CGAGCAAAAG ATAATGTCCC GAAAATTTAG CATTTAACTT TAAACATAAA 540
 AACGTCAGCT ATCACATGAC AGAAGACTAT GAACAGTTTC AATAATGTTT ATAGTAATCA 600
 TGTTAATAAC TGACGTTTAT TTTATTCTGC AGAATACTCT TCTAAATCTA TATTGCTGTG 660
 50 CCCATTTAAT GCTAAATCAG CAAATCGACC TTGCTGATAC AAATAGTGGC CGGCAACGCC 720
 TATCATTGCA GCATTATCTG TGCATAATTT AGGACTTGGG ATAGTTAATT GAATGTCATT 780

55

	AACAATTAAT CGCTGAACAC CATATTCTTT ACAAGCTTGA ATAGCTTTAA ACGTGAGCAC	900
	CTCTACAACA CTGTTTTGAA AGCTCGTTGC TACGTTAGCT TCAATGATTG GaATATTTTT	960
5	TTGTCGTTGA TTGTGAAGTT GATTGATTAC GGCACTTTTT AACCCTACTAA AACTAAAATC	1020
	ATAACTATCT TTATCCAACC AAACACGAGG GAATGAATAA GTATCTTCAC CTTCAGCAGC	1080
10	CAACCGATCA ACTTGTGGAC CACCTGGATA ATTAAACCA ATTGTTCTGT CCACTTTATC	1140
	ATAAGCCTCA CCTACTGCGT CATCTCGTGT TTCACCAATG ACTTCAAATG ATAAATGATC	1200
	CTTCATATAA ACTAATTCAG TATGTCCACC TGAAACAATA AGTGCAATTA GCGGGAATGT	1260
15	TAATGGCTCT TCTATGTGAT TAGCATATAT ATGTCCTGCA ATATGATGAA CAGGAATAAG	1320
	TGGCTnATCG TAAGCAAATG CCAATGCTTT GGCTGCATTA ACACCTATTA GTAACGCACC	1380
	AATTAGTCCA GGGCCTTCTG TAACCGCTAT GGCATCAATA TCTTCTATTG ATACATCGGC	1440
20	ATCCCCTAGA GCCTCGTTTA TTGTTGCTGT TATACCTTCA ACGTGATGTC TACTTGCCAC	1500
	TTCGGGAACG ACACCGCCAA ATCGTTTATG ACTTTCAATC TGACTTAAAA CTGTATTTGA	1560
	TAAAAATATCT CTGCCATTTT TTATAACACT AACGCTTGTT TCATCACAAC TTGTTTCAAC	1620
25	AGCTAGTATT AATATATCTT TAGTCATTTA AATTCACCCA CATAACCATT GCGTCCTCAC	1680
	CTTCACCATA ATAATTTTTA CGTTTACCAC CATATTGAAA TCCTAAATTT TCATATACAT	1740
	GTTGTGCCAC TTTATTATTA ACTCTTACTT CTAAACTCAT CACATCACAA GTGTGACTTG	1800
30	CATAGTTTAT TCCGTATTTT AAAAGCATTT GACCTAAACC ATAGCCTCTA TAATTATCAT	1860
	CGATTGCAAC TGTGTGAATT TGAGCTTGAT CGATAACAAT CCATAAACCT AAATAACCAA	1920
35	TAATTTGTTG TTCAAATTCt AAGACAAAAT ATTTGCGAAA GTTATTTTGC TCTATTTTCAT	1980
	GATAAAATGC GTCAATTGTC CAAGAACTGT CATTGAAACT CCGACGCTCA AGATCAAAGA	2040
	CTTGTGGCAC ATCTTCTTTA GTCATCTCTC TAATGTTTAA TTGTTCTTTT GACTGTTGAT	2100
40	CCAATTTTCGT TCCGCCTCAG CTAATTTATG GTATTTAGGA GTAAATGTAT GTACGTCTGA	2160
	AGGTTTATCT AGCAATTGAT ACATGACTGA TGCATTTGGT AGctGCGCAA TCACTTCACC	2220
	TTGTAATTCA TCTTGTAATT TTACAGTATC TTTCCCAATA TAAATAAATG GTTGGTTTAA	2280
45	ATCTTCTAAA AAAGCTCGCA ATGCCTCTAT CGACATATAT TGATCTTCTA AAATAGTCAC	2340
	TAATTGACCA TTTTGCCACT GGAATATGCC TGTATAAACT GCTTGTGCTC TTGCATCAAA	2400
	CACAGGAACC AATAATTTAT CAGTATGATC GATTGTTGCT GCCAATGCCT TTAATGATGA	2460
50	AACACCATAT AATTTAACAT CTAACGCATA CGCTAATGTT TTAGCAACAG TAACACCGAT	2520
	ACGTAAGCCA GTATATGAAC CAGGACCTTC AGCAACAATA ATCGCATCTA ATTGCTGTTT	2580

55

	TTGTTTAGAA TCCGTAGTTA TTTCAGCTAA AACTTCATCG TTTTGCATCA ATGCTACTGA	2700
	TAATGGTTGA TTCGATGTAT CAATGAGCAG CGAATTCATG GATAATTGCC TCCTTAATTT	2760
5	GTTCATAATG TTCTCCTTGC GCGAACAACCT CAATTTGTCT TGTATTTTCA GATATTGTTG	2820
	AAATGTTAAT AGATAAATGC GTCGCTGGAA GTAAATCTTT TATAAATTGA CTCCATTCAA	2880
	TAACAGTAAT TGCCTGATCT TCGAAAAATT CATCAAATCC TAAATCTTCA TCAGAATCTT	2940
10	CTAAGCGATA ACAATCCATA TGATGCAATT TTAAATTTTT ACCCCTATAT GATTTAATGA	3000
	TGTTAAATGT CGGGGAATTA ATCGTACGTC TTACACCAAG AGCTTTTCCT ATAAATTGCC	3060
	TTAACGTTGT TTTACCTGCT CCTAAATCTC CGTTAAGTAA AATCAAATCA CCACTTTTCA	3120
15	ATTGCTCAAC TAAAAATATA GCAAATTGAT TCATTTTCATC TAAATTATTT ATCTTTATCA	3180
	ATGTTGATTC TCCTATATTA TGCTTTTCAT TCATAAAAAAT GATTATCCAT TGTTCAATCG	3240
20	TATCTAACTT TATATTTAAC CTTTATATTG TAACAAATTT CAACTTAAAT TTCTTATCTT	3300
	TGAAACAGAT TATCTATTCA AAGTTAATTG TAAGAAAATT TAAATATTT GTTGACATAC	3360
	TAAAGCAGAT ATAGTAAATT AAATTTATCA AATTTTITAGA CAATTCTAAC TATTAAAGTG	3420
25	ATATATACCA TTCACGGAAG GAGTATAATA AAATGCTTAA TCAATATACT GAACATCAAC	3480
	CGACAACTTC AAATATTATT ATTTTATTAT ACTCTTTAGG ACTCGAACGT TAgtAAATAT	3540
	TTACTAAACG CTTTAAGTCC TATTTCTGTT TGAATGGGAC TTGTAAACGT CCCAATAATA	3600
30	TTGGGACGTT TTTTTATGTT TTATCTTTCA ATTACTTATT TTTATTACTA TAAAACATGA	3660
	TTAATCATTa AAATTTACGG GGGAATTTAC TATGCGAaCG AgcATGATCA AAAAAGGAGA	3720
	TCACCAAGCA CCAGCAAGAA GTCTTTTACA TGCCACGGGC GCGCTAAAAA GTCCAATGA	3780
35	TATGAACAAA CCATTTGTAG CTATTTGTAA CTCTTATATT GATATTGTTT CTGGACATGT	3840
	TCaCTTGAGA GAGCTTGcAG ATATAGCTAA AGAAGCAATT AGAGAAGCCG GTGCCATTCC	3900
40	ATTTGAATTC AATACAATTG GTGTTGATGA TGGAATAGCT ATGGGACATA TCGGAATGCG	3960
	ATATTCTCTA CCATCACGTG AAATTATTGC AGATGCAGCT GAAACTGTAA TTAACGCTCA	4020
	TTGGTTTGAC GCGTATTTT ACATTCCTAA TTGTGACAAG ATTACACCCG GTATGATTTT	4080
45	AGCAGCCATG AGGACAAACG TACCAGCTAT CTTTGTCTCT GGTGGACCAA TGAAAGCTGG	4140
	CTTATCTGCA CATGGAAAAG CATTAACTT TTCATCAATG TTTGAAGCAG TCGGCGCATT	4200
	TAAAGAAGGA TCGATTCTA AAGAAGAATT TTTAGATATG GAACAAAATG CCTGCCCTAC	4260
50	TTGTGGTTCA TGTGCTGGGA TGTTTACTGC AAATTCAATG AACTGTTTGA TGGAAGTTTT	4320
	AGGTCTAGCA TTACCATACA ACGGTACTGC ACTTGcAGTC AGTGATCAGC GACGAGAAAT	4380

55

	TATCGTTACT	CGCGAAgCAA	TTGATGATGC	ATTTGCACTT	GATATGGCTA	TGGGTGGTTC	4500
	AACAAACACG	GTA CTGCATA	CGTTAGCCAT	TGCCAATGAA	GCTGGTATTG	ATTATGACTT	4560
5	AGAGCGCATT	AATGCTATTG	CCAAACGCAC	GCCATATTTA	TCAAAAATAG	CACCTAGTTC	4620
	ATCGTATTCA	ATGCATGATG	TGCATGAAGC	TGGTGGCGTC	CCAGCAATTA	TTAATGAATT	4680
	GATGAAGAAA	GATGGCACGT	TACACCCAGA	TAGAATCACA	GTTACTGGCA	AAACGTTACG	4740
10	TGAAAATAAC	GAAGGCAAAG	AAATTAAGAA	CTTTGATGTC	ATTCACCCTC	TTGATGCACC	4800
	ATATGATGCA	CAAGGCGGTT	TATCTATCTT	ATTTGGTAAT	ATCGCCCCTA	AAGGCGCAGT	4860
	TATTAAAGTT	GGCGGCGTTG	ATCCATCTAT	CAAAACATTT	ACTGGGAAAG	CAATTTGTTT	4920
15	CAATTCGCAT	GATGAAGCTG	TTGAAGCAAT	AGACAATCGT	ACCGTTCGTG	CAGGCCACGT	4980
	CGTTGTCATT	AGATATGAAG	GACCTAAAGG	TGGACCAGGT	ATGCCTGAAA	TGTTAGCACC	5040
20	TACTTCCTCT	ATTGTTGGTC	GCGGCTTAGG	TAAAGATGTT	GCATTAATTA	CTGATGGGCG	5100
	TTTTTCCGGT	GCCACAAGAG	GTATTGCAGT	TGGTCATATT	TCCCCTGAAG	CTGCATCTGG	5160
	TGGACCAATT	GCCTTAATTG	AAGATGGTGA	TGAGATTACT	ATTGATTTAA	CAAATCGTAC	5220
25	ATTAAACGTA	AACCAGCCTG	AAGATGTTCT	AGCGCGTCGC	CGAGAATCTT	TAACACCATT	5280
	TAAAGCGAAA	GTAAAAACAG	GTTATCTAGC	TCGTTATACT	GCCCTAGTAA	CTAGCGCAAA	5340
	TACAGGTGGC	GTCATGCAAG	TCCCTGAGAA	TTTAATTTAA	TTTATTTTTA	TATTGGAGAT	5400
30	GGTTAAAATG	TCTAAAACTC	AACATGAAGT	AAACCAAAAT	ATTGACCCTT	TAAAAATGGC	5460
	TGAATCACTT	GAACCTGAAC	AAC TAAATGA	AAAAACTTTA	AATGATATGC	G TTCAGGATC	5520
	AGAAGTGCTA	GTAGAAGCTC	TACTTAAAGA	AAATGTGGAT	TATTTATTCG	GTTATCCTGG	5580
35	TGGTGCCGTA	CTACCTTTAT	ATGACACGTT	TTATGATGGT	AAAATCAAAC	ATATTTTAGC	5640
	AAGA CACGAA	CAAGGTGCTG	TTCATGCTGC	AGAAGGTTAT	GCACGTGTAT	CTGGTAAmT	5700
40	GGCGTCGTTG	TAGTTACAAG	CGGTCCaGGT	GCAACTAATG	TAATGACAGG	TATTACGGAT	5760
	GCACATTGCG	ACTCTTTACC	TCTAGTTGTA	T TCACTGGAC	AAGTTGCTAC	ACCAGGCATT	5820
	GGTAAAGATG	CATTCCAAGA	AGCGGATATT	CTATCTATGA	CTTCACCAAT	TACAAAACAA	5880
45	AATTATCAAG	TGAAACGTGT	TGAAGATATC	CCTAAAATCG	TACACGAAGC	TTTCCATGTA	5940
	GCTAATTCTG	GACGCAAAGG	TCCTGTAGTG	ATTGATTTTC	CAAAAGATAT	GGGTGTTTTA	6000
	GCTACAAATG	TGGATTTATG	CGACGAAATC	AATATTCCAG	GTTATGAAGT	TGTTACAGAA	6060
50	CCAGAAAATA	AAGACATTGA	CACTTTCATC	TCACTTTTAA	AAGAAGCGAA	AAAGCCTGTC	6120
	GTATTAGCCG	GCGCAGGTAT	TAATCAATCA	AAATCAAATC	AATTATTAAC	ACAGTTTGTT	6180
55							

	GATACACTAT TTTTAGGTAT GGGAGGAATG CATGGTTCTT ATGCTAGTAA CATGGCATT	6300
	ACTGAGTGTG ATTTACTCAT TAATTTAGGT AGCCGCTTCG ATGATAGATT AGCAAGCAAA	6360
5	CCTGATGCCT TTGCACCTAA CGCCAAAATT GTACATGTAG ATATTGATCC TTCAGAAATC	6420
	AATAAAGTTA TTCATGTAGA TTTAGGTATT ATTGCAGACT GTAAAAGATT TTTAGAATGT	6480
10	TTAAATGATA AAAATGTTGA GACTATAGAA CACAGTGA CT GGGTTAAACA TTGTCAAAT	6540
	AATAAGCAGA AACACCCATT TAAACTTGGT GAAGAAGATC AAGTATTTTG TAAGCCACAA	6600
	CAAACAATCG AATATATCGG CAAAATTACA AATGGTGAAG CAATTGTTAC TACAGACGTG	6660
15	GGACAACATC AAATGTGGGC AGCTCAATTT TATCCATTTA AAAATCACGG ACAATGGGTT	6720
	ACAAGCGGTG GTTTAGGAAC AATGGGATTC GGTATTCCTT CGTCAATTGG TGCCAAATTA	6780
	GCTAATCCTG ATAAAACAGT CGTATGTTTC GTCGGTGACG GTGGTTTCCA AATGACAAAC	6840
20	CAAGAAATGG CACTTTTACC CGAATATGGT TTAGATGTCA AAATCGTACT AATCAATAAT	6900
	CGAACATTAG GTATGGTTAA ACAATGGCAA GATAAGTTCT TTAATCAACG CTTCTCACAC	6960
	TCAGTATTTA ATGGTCAACC TGATTTTATG AAAATGGCAG AAGCATATGG CGTCAAAGGT	7020
25	TTCTTAATCG ATAAGCCAGA ACAACTGGAA GAACAATTAG ATGCAGCGTT TGCTTATCAA	7080
	GGACCAGCTT TAATTGAGGT TCGTATTTCC CTTACTGAAG CTGTAACCCC AATGGTCCG	7140
	AGTGGCAAAT CAAATCATGA AATGGAGGGC TTATAATGAC AAGAATTCTT AAATTACAAG	7200
30	TTGCGGATCA AGTCAGCAGC CTAAATCGAA TTACAAGTGC TTTTGTTCGC CTACAATATA	7260
	ATATCGATAC ATTACATGTT ACACATTCTG AACACCTGG GATTTCTAAC ATGGAAATTC	7320
35	AAGTCGATAT TCAAGATGAT ACATCACTTC ATATATTAAT TAAAAAATTA AAACAACAAA	7380
	TTAATGTTTT AACGGTTGAA TGCTACGACC TTGTTGATAA CGAAGCTTAA TTTTAAGACA	7440
	AAGGCAATGA TGCGCTAATT AGTTATAGAT ATATCATAGG CTGCTAGTTA ACATCTGCCA	7500
40	CTATTACAAA GTTATATTTT AGAATTTTCG AAACACAAAA TATTTAATTA TTTGGAGGAA	7560
	TTTATTATGA CAACAGTTTA TTATGATCAA GATGTAAAAA CGGACGCTTT ACAAGGCAAA	7620
	AAAATTGCAG TAGTAGGTTA TGGATCACAA GGTCACGCGC ATGCACAAAA CTTAAAAGAC	7680
45	AATGGATATG ATGTAGTCAT CGGCATTTCG CCAGGTCGTT CTTTGTGACAA AGCTAAAGAA	7740
	GATGGATTTG ATGTGTTCCC TGTTGCAGAA GCAGTTAAGC AAGCTGATGT AATTATGGTG	7800
	CTATTACCTG ATGAAATTCA AGGTGATGTA TACAAAAACG AAATTGAACC AAATTTAGAA	7860
50	AAACATAATG CGCTTGCAAT TGCTCATGGC TTAAACATTC ATTTTGGTGT TATTCAACCA	7920
	CCAGCTGATG TTGATGTATT TTTAGTAGCT CCTAAAGGAC CGGGTCATTT AGTTAGACGT	7980

55

	CAAGCACGTA ATATTGCTTT AAGTTATGCA AAAGGTATTG GTGCAaCTCG TGCAGGTGTT	8100
	ATTGAAACAA CATTTAAGA AGAAACTGAG ACAGATTTAT TTGGTGAACA AGCAGTACTT	8160
5	TGCGGTGGTG TATCGAAATT AATTCAAAGT GGCTTTGAAA CATTAGTAGA AGCGGGTTAT	8220
	CAACCAGAAT TAGCTTATTT TGAAGTATTA CATGAAATGA AATTAATCGT TGATTTGATG	8280
	TATGAAGGCG GTATGGAAAA TGTACGTTAC TCAATTTCAA ATACTGCTGA ATTTGGTGAC	8340
10	TATGTTTCAG GACCACGTGT TATCACACCA GATGTTAAAG AAAATATGAA AGCTGTATTA	8400
	ACTGATATCC AAAATGGTAA CTTCACTAAT CGCTTTATCG AAGACAATAA AAATGGATTC	8460
	AAAGAATTTT ATAAATTACG CGAAGAACAA CATGGTCATC AAATTGAAAA AGTTGGTCGT	8520
15	GAATTACGCG AAATGATGCC TTTTATTAAA TCTAAAAGCA TTGAAAAATA AGATAGACCT	8580
	ACAATGAGGA GTTGTAAAT ATGAGTAGTC ATATTCAAAT TTTTGATACG AACTAAGAG	8640
20	ACGGTGaACA AACACCAGGA GTGAATTTTA CTTTGTATGA ACGCTTGCGT ATTGCATTGC	8700
	AATTAGAAAA ATGGGGTGTA GATGTTATTG AAGCTGGATT TCCTGCTTCA AGTACAGGTA	8760
	GCTTTAAATC TGTTCAAGCA ATTGCACAAA CATTAAACAC AACGGCTGTA TGTGGTTTAG	8820
25	CTAGATGTAA AAAATCTGAC ATCGATGCTG TATATGAAGC AACAAAAGAT GCAGCGAAgC	8880
	CGGTcGTGCA TGTTTTTATA GCAACATCAC CTATTCATCT TGAACATAAA CTTAAAATGT	8940
	CTCAAGAAGA CGTTTTAGCA TCTATTAAAG AACATGTCAC ATACGCGAAA CAATTATTTG	9000
30	ACGTTGTTCA ATTTTCACCT GAAGATGCAA CGCGTACTGA ATTACCATTG TTAGTGAAAT	9060
	GTGTACAAAC TGCCGTTGAC GCTGGAGCTA CAGTTATTAA TATTCCTGAT ACAGTCGGCT	9120
	ACAGTTACCA TGATGAATAT GCACATATTT TCAAAACCTT AACAGAATCT GTAACATCTT	9180
35	CAAATGAAAT TATTTATAGT GCTCATTGCC ATGACGATTT AGGAATGGCT GTTTCAAATA	9240
	GTTTAgCTGC AATTGAAGGC GGTGCGAGAC GAATTGAAGG CACTGTAAAT GGTATTGGTG	9300
40	AACGAGCAGG TAATGCAGCA CTTGAAGAAG TCGCGCTTGC ACTATACGTT CGAAATGATC	9360
	ATTATGGTGC TCAAACCTGCT CTTAATCTCG AAGAACTAA AAAACATCG GATTTAATTT	9420
	CAAGATATGC AGGTATTCGA GTGCCTAGAA ATAAAGCAAT TGTGGCCAA AATGCATTTA	9480
45	GTCATGAATC AGGTATTCAC CAAGATGGCG TATTAAAACA TCGTGAAACA TATGAAATTA	9540
	TGACACCTCA ACTTGTTGGT GTAAGCACGA CTGAACTTCC ATTAGGAAAA TTATCTGGTA	9600
	AACACGCCTT CTCAGAGAAG TTAAGCAT TAGGTTATGA CATTGATAAA GAAGCGCAA	9660
50	TAGATTTATT TAAACAATTC AAGGCCATTG CGGACAAAAA GAAATCTGTT TCAGATAGAG	9720
	ATATTCATGC GATTATTCAA GGTTCTGAGC ATGAGCATCA AGCACTTTAT AAATTGAAA	9780

55

	AAGAGGGTCA TATTTACCAG GATTCAAGTA TTGGTACTGG TTCAATCGTA GCAATTTACA	9900
	ATGCAGTTGA TCGTATTTTC CAGAAAGAAA CAGAATTAAT TGATTATCGT ATTAATTCTG	9960
5	TCACTGAAGG TACTGATGCC CAAGCAGAAG TACATGTAAA TTTATTGATT GAAGGTAAGA	10020
	CTGTCAATGG CTTTGGTATT GATCATGATA TTTTACAAGC CTCTTGTAAG GCATACGTAG	10080
	AAGCACATGC TAAATTTGCA GCTGAAAATG TTGAGAAGGT AGGTAATTAA TTATGACTTA	10140
10	TAACATTGTT GCCCTACCTG GTGATGGAAT CGGTCCAGAA ATTTTGAACG GATCTCTATC	10200
	ATTGCTTGAA ATTATAAGTA ATAAATATAA CTTTAATTAT CAAATAGAGC ACCACGAATT	10260
	TGGTGGTGCC TCTATTGATA CATTCGGCGA GCCTTTAACT GAGAAAACCT TAAATGCGTG	10320
15	TAAAAGAGCA GATGCTATTT TACTGGGTGC AATCGGTGGA CCTAAATGGA CAGATCCTAA	10380
	CAATCGACCA GAACAAGGAT TATTAATAAT GCGTAAATCC TTAAATTTAT TTGTAAATAT	10440
20	ACGCCCACT ACCGTTGTCA AAGGCGCTAG TTCTTTATCA CCTTTAAAGG AAGAACGCGT	10500
	TGAAGGCACA GATTTAGTTA TAGTCCGTGA ATTGACAAGT GGTATTTATT TTGGAGAACC	10560
	TAGACATTTT AATAATCAG AGGCCTTAGA TTCTCTTACT TATACAAGAG AAGAAATAGA	10620
25	ACGCATTGTT CACGTAGCAT TTAAATTGGC CGCTTCAAGA CGAGGAAAAC TAACATCAGT	10680
	TGATAAAGAA AATGTATTAG CTTCTAGTAA ATTGTGGCGC AAAGTCGTAA ATGAAGTAAG	10740
	TCAATTATAT CCAGAAGTAA CAGTAAATCA CTTATTTGTT GATGCTTGTA GTATGCATTT	10800
30	AATCACAAAT CCAAACAAT TTGACGTCAT CGTATGTGAA AACTTATTTG GCGATATTTT	10860
	AAGTGATGAA GCTTCAGTGA TTCCTGGTTC ACTTGGTTTA TCACCTTCTG CTAGTTTTAG	10920
	TAACGATGGT CCAAGATTGT ATGAGCCTAT TCATGGATCA GCACCAGATA TTGCAGGTAA	10980
35	AAACGTTGCC AATCCATTG GAATGATTCT ATCTTTAGCG ATGTGTTTAC GTGAAAGCTT	11040
	AAATCAACCA GATGCTGCAG ATGAATTAGA ACAACATATT TATAGCATGA TTGAACATGG	11100
40	GCAAACGACA GCAGATTTAG GCGGCAAATT GAATACTACT GATATTTTCG AAATTCTATC	11160
	TCAAAAATTG AATCACTAAG GGGGAGATGT AAATGGGTCA AACATTATTT GACAAGGTGT	11220
	GGAACAGACA TGTGTTATAC GGGAAATTGG GCGAACCACA ACTATTATAC ATTGATTTAC	11280
45	ACCTTATACA TGAAGTTACT TCTCCTCAAG CATTTGAAGG ACTTAGGCTT CAAAACAGAA	11340
	AATTAAGACG CCCAGATTTA ACATTTGCAA CACTCGATCA CAATGTTCTT ACTATTGATA	11400
	TATTCAATAT TAAAGATGAA ATTGCAAACA AACAAATCAC AACATTACAA AAAAACGCCA	11460
50	TAGATTTTGG GGTGCATATT TTTGATATGG GTTCTGATGA ACAAGGTATT GTTCACATGG	11520
	TAGGACCTGA GACAGGACTT ACACAGCCTG GCAAGACAAT CGTTTGTGGT GACTCTCACA	11580

55

	ATGTTTTTCGC AACTCAAACG CTATGGCAAA CAAAACCCAA AACTTAAAA ATCGATATTA	11700
	ATGGTACCTT ACCAACAGGC GTCTATGCTA AGGACATTAT TCTGCATTTA ATTAAAAACGT	11760
5	ATGGTGTGTA CTTTGGTACA GGCTATGCTT TGGAAATTAC TGGCGAAACA ATTAAAAACC	11820
	TTTCAATGGA TGGTCGAATG ACTATTTGTA ACATGGCTAT CGAAGGTGGT GCCAAATACG	11880
	GCATAATCCA ACCTGATGAT ATAACATTTG AATATGTTAA AGGGAGACCA TTTGCCGATA	11940
10	ACTtCGCTAA ATCAGTTGAT AAGTGGCGTG AgCTATATTC TGATGACGAC GCGATATTTG	12000
	ATCGTGAAT TGAAC TTGAT GTTTCAACAT TAGAACCACA AGTGACATGG GGAAC TAATC	12060
	CTGAAATGGG TGTTAATTTT AGTGAACCAT TCCCTGAAAT CAATGATATC AACGATCAAC	12120
15	GTGCGTATGA TTATATGGGG TTAGAACCAG GTCAAAAAGC TGAAGACATC GACTTAGGGT	12180
	ATGTTTTTCT CGGTTCATGT ACAAATGCTA GACTATCAGA TTTGATTGAA GCTAGTCATA	12240
20	TTGTTAAAGG AAATAAAGTT CATCCAAATA TTACAGCTAT TGTCGTACCA GGTTCCTCGTA	12300
	CAGTAAAAAA AGAAGCAGAA AAATTAGGTC TAGATACTAT CTTTAAAAAT GCAGGATTTG	12360
	AATGGCGTGA ACCAGGATGT TCAATGTGTT TAGGCATGAA TCCTGACCAA GTACCTGAGG	12420
25	GCGTACATTG TGCATCTACA AGTAATCGAA ACTTTGAAGG ACGACAAGGC AAAGGTGCAA	12480
	GAACACATTT AGTATCCCCT GCTATGGCAG CAGCAGCAGC TATTCATGGT AAATTTGTGG	12540
	ACGTAAGAAA GGTGGTTGTT TAAATGGCAG CAATCAAACC TATTACAACA TATAAAGGTA	12600
30	AAATAGTCCC TCTCTCAAC GACAATATCG ATACAGACCA AATCATTCCT AAGGTACACT	12660
	TAAAGCGTAT TTCAAAAAGT GGCTTTGGTC CATTTGCTTT TGATGAATGG CGGTACTTAC	12720
	CTGATGGTTC AGATAATCCT GATTTCAATC CTAACAAACC ACAATATAAA GGGGCTTCTA	12780
35	TTTTAATTAC TGGAGATAAT TTGGATGTG GTTCAAGTCG TGAACATGCT GCTTGGGCTC	12840
	TTAAGGACTA TGGTTTTCAT ATTATTATTG CAGGAAGTTT CAGTGACATA TTTTATATGA	12900
40	ATTGCACTAA AAATGCGATG TTGCCTATCG TTTTAGAAAA AAGTGCCCGT GAACATCTTG	12960
	CACAAATATGT TGAAATTGAG GTCGATTTAC CAAATCAAAC TGTGTCATCA CCAGACAAGC	13020
	GTTTCCATTT TGAAATTGAT GAAACTTGGA AGAATAAACT TGTAATGGC TTAGATGACA	13080
45	TTGCAATCAC CCTACAATAT GAATCATTAA TAGAAAAATA TGAAAAATCa CTTTAAGGGA	13140
	GTTGAATATT ATGACAGTCA AAACAACAGT TTCTACGAAA GATATCGATG AGGCATTTTTT	13200
	AAGACTTAAA GATATTGTCA AAGAAACACC TTTACAATTA GACCATTACT TATCTCAAAA	13260
50	GTATGATTGT AAAGTCTATT TAAAACGAGA AGATTTIACAA TGGGTACGTT CTTTTAAATT	13320
	AAGAGGTGCT TACAACGCTA TTTCTGTTTT ATCAGATGAA GCTAAAAGTA AAGGTATTAC	13380

55

	AAACGCTGTT ATCTTTATGC CAGTCACTAC ACCTTTACAA AAGGTAAATC AAGTAAAGTT	13500
	CTTTGGAAAT AGTAACGTTG AAGTTGTACT CACTGGTGAT ACATTTGATC ACTGTTTAGC	13560
5	TGAAGCTTTA ACTTATACAA GTGAACATCA AATGAACTTT ATAGATCCAT TCAATAATGT	13620
	TCATACAATT TCTGGACAAG GTACGCTTGC TAAAGAAATG CTAGAACAAG CAAAGTCTGA	13680
	CAATGTTAAC TTTGATTATC TATTTGCCGC AATTGGTGGT GGCGGTTTAA TTTCAGGTAT	13740
10	TAGTACTTAC TTTAAACCT ATTACCTAC CACGAAAATT ATAGGTGTTG AACCTTCAGG	13800
	TGCAAGTAGT ATGTATGAAT CTGTTGTGGT AAATAATCAG GTAGTCACAT TGCCTAATAT	13860
	CGATAAATTT GTGGACGGTG CATCTGTAGC TAGAGTTGGC GATATTACAT TTGAAATTCG	13920
15	AAAAGAAAAT GTAGATGATT ACGTTCAAGT AGATGAAGGT GCAGTTTGTT CTACGATTTT	13980
	AGATATGTAT TCAAAACAAG CAATTGTAGC AGAACCTGCT GGCGCATTAA GTGTAAGTGC	14040
20	GCTTGAAAAC TATAAAGATC ATATTAAAGG TAAAACAGTG GTTTGTGTCA TTAGTGGTGG	14100
	TAATAATGAT ATTAATCGAA TGAAAGAAAT TGAAGAACGT TCATTACTAT ACGAAGAAAT	14160
	GAAGCATTAC TTTATCTTAA ATTTCCCTCA ACGTCCAGGT GCATTGAGAG AATTTGTAAA	14220
25	TGACGTATTA GGACCTCAAG ACGATATTAC TAAATTTGAA TACTTAAAAA AATCTTCTCA	14280
	AAATACAGGT ACTGTCATTA TTGGTATTCA ACTTAAAGAT CATGATGATT TAATACAACT	14340
	CAAACAACGT GTAAATCATT TCGATCCTTC CAATATTTAT ATTAATGAAA ATAAGATGTT	14400
30	ATATTCATTG TTAATTTAAC ACATAGTAAG AAAACAGTC ATAAATTGAT TTCTAATTGA	14460
	AATCATCTTA TGACTGCTTT TTATTATACT TTACATTTCT CGTTTCGTCA GATTCAAACG	14520
	TTTTCACTTC GCCAAGCCAT CTTTCTTTGT GTTTGCTTTT aTTTGGACGT TTTAGACATA	14580
35	AAAAAaGAGA CTTTGCGGTC TCAATGCGGC TCATCGCATC CACTTTTTGC CTGGCAACGT	14640
	TCTACTCTAG CGGAACGTAA GTTCGaCTAC CATCGACGCT AAGGAGCTTA ACTTCTGTGT	14700
40	TCGGCATGGG AACAGGTGTG ACCTCCTTGC TATAGTCACC AGACATATGA ATGTAATTTA	14760
	TACATTCAAA ACTAGATAGT AAGTAAAAGT GATTTTGCTT CGCAAAACAT TTATTTTGAT	14820
	TAAGTCTTCG ATCGATTAGT ATTCGTCAGC TCCACATGTC ACCATGCTTC CACCTCGAAC	14880
45	CTATTAACCT CATCATCTTT GAGGGATCTT ATAACCGAAG TTGGGAAATC TCATCTTGAG	14940
	GGGGGCTTCA TGCTTAGATG CTTTCAGCAC TTATCCCGTC CACACATAGC TACCCAGCTA	15000
	TGCCGTTGGC ACGACAACCTG GTACACCAGA GGTATGTCCA TCCCGGTCCT CTCGTACTAA	15060
50	GGACAGCTCC TCTCAAATTT CCTACGCCCA CGACGGATAG GGACCGAACT GTCTCACGAC	15120
	GTCTGAACC CAGCTCGCGT ACCGCTTTaA TGGGCGAACA GCCCAACCCT TGGGACCGAC	15180

55

GTGGAACTT

15249

(2) INFORMATION FOR SEQ ID NO: 103:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14051 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

15	GTGGCAATAT TTCTAGTTCT CGTTTGTATA AGATTTTAAA AGGATCTGTT GTGTTTGCAG	60
	TGTCCTGATT TGAATTAGAT ACAAATTCAT TCACTAAAGA TGTGTAAAGT TTCATATCTA	120
	CATATGTTTC ACCTTTATAT ACAGTTCGAA TAGCTAACAA TAATTGTTCA TCAGGTGCAT	180
20	TTTTCAATAT GTAACCTTTC GCACCATTAC GCAACACATG GAACAAATAC TCCTCATCAT	240
	CAAACATTGT TAATATTAGT ATTTTAGTTT CAGGAAAAC TGCAGCAATT TTA CTCTAG	300
	CGATAAGACC TGACTCACCT GGTGGcATAC TTAAATCCAT TAGTAACACA TCAGGTTTAt	360
25	ATTCCATTAC TTTTGGTAA GCTTCGACGC CATCTGCAGC CGTTGCAACA ACTTCCATAT	420
	CATTTTGTATA ATTTAAAATC ATAGAGAACC CCGTACGGAC AACAGCGTGA TCATCGGCAA	480
	TGACTATTTT CAATTTTATT CCCCCAATGT ATGTTTCAA TTGGAATGTT CAATGTAACA	540
30	TTGGTACCCT CACCAATTTT CGTTTCAATA TTGACGCTAC CGCTGACTAA CTCAGCTCGC	600
	TCATTCATT C CATATAAACC GAGTCCAGAA CCTTTAGGCT TAGAACTTGG ATCAAAACCA	660
	TTTCCCGCAT CTATCACTTC TGCTACCAA TGGCGCCCAG TTTGACGGAT ACCTACATTT	720
35	ATTTCAATTA CATCAGCGTA TTTCAACGCA TTTAAAATAG CTTCTTGAC TACTCGATAA	780
	ACAACCGTTT CAATATCACT ATCAAAGCGA GTATTTTAA TATTTGATGT ATATATGATT	840
40	TTTATTCCAT AATTTTCTTC AAAGTGT TTA AAATATGATT TAAAAGCTGC TTCAAGGCCT	900
	AGATCATCCA AAGAAGCGGG TCTTAATTCA ACCGACATAT TACGTATATC ATCAATTAAT	960
	TTAGCGACAA TATATTCAAT ATTTTCTGCG TCTTCCAAAA GCTTAGTTGT ATCTTCTTGA	1020
45	TATTTTAATA ATCTCAATTG AACATCTACA TTGAGCATTT CTTGAATCAC ACTATCATGT	1080
	AACTCTCTAG AAATTCGCTT TCTTTCAATT TCTTGGGCTG AGATTGTTTT ACGCATCATA	1140
	CGTTGTTGAT GCAATTTCTC TTGCTGTTCA ATTTGTGATG AAACATTTTG AAGCGTAAAT	1200
50	GCATGAATTC CCCTGTCTTG ATCAATCAAC TGATATGTTG CTGTAAATGG CATCACTTTT	1260
	TGATCTTTTC TCTTCATAAA TACTTGGAAA TTCGTAGCTT GTACTTG CAT CGATTCTAAG	1320

55

	ATCGCATTCTG CCACAGCACT GTAATTATCT TCTTCAGATA ATATATCTTT AGCAGCATCA	1440
	TTTCATTGCAA TAATTTTACC GTTATCATCA GCAAAAACTA TCTTTTCGAT TGAATGCTCA	1500
5	TAATATTTTT TCAATAAAGT ATCTAACTGT ATACTGTCCT CATTAATCAT GACTTACACC	1560
	CTAATTCATC TCATTATTTA TCATCATTGA AAATACCAAA CTTACGTTGA ATATCATCAT	1620
	TATCAAATAT TTTTGGTAAA GGACGACCAT CTCTTTGACC AAATAATAGT ACGCCATACA	1680
10	CTTGATTCTT ATACCAAAGC GGCACGCTA AAAGTCTGT TAATGATTCTG CTCAATAAAA	1740
	TTGGATAGTC AATCTTTTCT TCAGGCCCTA AAGCTAAACC AACATTGGCT ATTACCATAC	1800
	GCTTCTCTGT TTTCATAACA GTTCCAGCTA ATCCACGACC TTTTCTTAAA ATAATCAATT	1860
15	TAAATCGATT ATTTTATTAT CCTGAAACAT AGTGCCATTT TATTGGAGAT GATGGTTTGT	1920
	TAGATTCTATA GAAAGCGATT GCCGCAAAAT CATAACCCTC TTCTTTGCGT ATTTTATCTA	1980
20	ATGTCTCTTG AAATCTACGA TCTTCAATTA TTGCTTCTGG TGTCAAATCC TTTCACCTCT	2040
	TATGCTTACA CTTTATTCTT ACGGTAAATA ATATATCTGC GATTTATATA TGTCAAAGGT	2100
	ACACTCCAAA CATGCACCAA ACGTGTAAT GGCCAACAAG CCATAATAGT GAAACCTAAC	2160
25	AATATATGCA TTTTAAATGC AATCGGCACA CCACTCATCA ATGACGCATC TGGTTTTAAC	2220
	ATAAATAATT GTCTAAACCA AATTGATAAT GAAGTTCTGT AGTTAAAGTC TGGATGTTGT	2280
	ATATTTGTTA CTAATGTTGC GTAACATCCC ATAAATACGA TAAGTAATAA TAAGAAATTT	2340
30	ACAAATATAT CCGACGCTGA ACTTAATCTT CGAATACTTT TCGTAGTAAC ACGTCTCGCT	2400
	GTTAATAAAA ACATCCCTAT CAAAGTTATT ATACCAAAGA TGCTACCAAT ATAAACAGCG	2460
35	CCTATATGAT ATAAATGCTC AGACACACCC ACTGCATCCA TCCATGGTTT CGGTATTAAAC	2520
	AATCCAACTA CGTGTCCAAA AAACACTGGA ATAATACCTA AGTGAAATAA TAAACTTCCC	2580
	CACATCAACC TTTTCTTTT TATTAATTCA CTAGATTTAG CTGTCCAAGA AAATTTATCA	2640
40	TAACGATAAC GTGCAATATG ACCTGCGACA AAGACAATA AACATAAATA CGGAAATATA	2700
	ACCCATAAAA ACTGATTAAG CATGATGTTT CACTCCTTTT GGTGATGTCA AACATAATTT	2760
	CAATGTTTTT CTAAGTGCTT GAATCACATA GGCATATGGA TTGTTATCTT CACCAAGTGC	2820
45	ATTCGCCATC ACATATGTTT CATCCTCAAT AATCATAATG ATTAATTGAA TATTCTCTTC	2880
	AGCTCTTGGA TCATTTGCGC ATTCTGCCAC TTGCAAAAAT TGAAGCATCA ACGGTAGATA	2940
	ATCAGAAAGT TCATTATCTA CCATTTCTAG TCCAAACATT TCATATAATA CCTTTAATTT	3000
50	AGCTAACATT TGCCACGTT CTTTTTTCGT ATCAAATTTG TTATACGTCA TATATAATGG	3060
	TGCTTTTTTC GTAAAATCAA ATGTATCTGT ATAAATCGCT TTGATTTCTG ATAATGAAAA	3120

55

	TGTTTCTTCA AAAGTTTTTG GATGAAAAGT TAATTTTCTT GGAAAACATA ACTGTTGTGC	3240
	CATATATCCA AAACTTTCTT GATATTTTTT AAAATTATCG AAATTAATCA CGGAAAATCC	3300
5	CTCCATAGAA ATTCTCATTA TAAATTTCTT GACCAGTTTT CCCTGAACCT ACTGCAACGC	3360
	CACAGCCTTC ACAGTTATCT CCAAAATGCT CGCCGCCGTA ATTGTATCCT GTACTACCTT	3420
10	GTGCGTGATA CGTATCTAAA TAGGTTTCTT TGTGTGATGT TGGAATAACA AATCGATCTT	3480
	CATATTTGGC TAGTCCTAAT AAACGATACA TGTCTTTAGT TTGGCGCTCG GTTATACCTA	3540
	ATCGCTCTAA TCGAGACGTG TCAAATGGCT GTTGAGTAAC TTGAGATCTC ATATAACTTC	3600
15	TCATCATTGC CATACTTGT AGGGCTCCTT TTACTGGCTC TGTATCTCCT GCAGTGAAAA	3660
	TATTAGCTAA GTATTCAATA GGTAAACGCA TTTCTTCAAT GGCTGGGAAA ATCGCATCTG	3720
	GATTTTGAGT TGTATTTTAA CCTTCAAAAT AGCTCATAAT TGGGCTAAGT GGTGGGCAAT	3780
20	ACCAAACCAT CGGCATCGTT CTAAATTCAG GATGTAACGG AAATGCAAGT TTATATTCAA	3840
	TTGCTAACTT ATAAATTGGA GAGTTTGTG CAGCTTCAAT CCAATCGTAA CCAATACCAT	3900
	CTTTTTCAGC TTGAGCAATG ACTTCTTCGT CAAATGGGTT TAAGAATATA TCTAATTGTT	3960
25	TTTCATATAA ATCTTTCTCG TCTACTGCTG AAGCTGCTTC ATGAACTCGA TCTGCATCAT	4020
	ATAATAAAAC ACCTAAGTAA CGCATACGTC CTGTACAAGT TTCAGAGCAT ACCGTAGGCA	4080
	TACCCGCCTC GATTCTCGGG AAACAGAAAG TACACTTTTC AGCTTTGTTC GTTTTCCAAT	4140
30	TGAAGTAAAC TTTCTTATAT GGACAACCTG TCATACAGTA ACGCCATCCA CGACATGCGT	4200
	CTTGGTCAAC TAATACAATG CCATCTTCAT CACGTTTATA CATAGCACCT GAAGGACACG	4260
35	ATGCAACGCA ACTTGGATTG AAGCAATGTT CACATAAAGC TGGTAAATAC ATCATAAAAG	4320
	TTTCGTCAAA TTGGAATTTA ATATCTTCTT CTATTTTTTG GATGTTAGGA TCTTTTGGAC	4380
	CTGTAAACATG ACCACCTGCT AAGTCATCTT CCCAGTTAGG TCCCCATTCA ATTTCAATGT	4440
40	TATCCCCCGT AATTTCTGAA TACGCTCTAG CAACTGGCGA ATGCTTCCCT GATTTGCGAG	4500
	TTGTTAAATG TTCATAATTA TAGTTCCATG GCTCATAATA ATCTTTAATT AATGGCATAT	4560
	CTGGGTTATA AAAAATTTTA CCTAAAGCAA TTTTGTAAAT TCTACTTCCA GATTTTAATT	4620
45	CAAGTTTCCC TTTACGATTT AGTACCCAAC CACCTTTGTA GTGTTCTTGG TCTTCCCAAC	4680
	GTTTCGGATA CCCTACACCT GGCTTCGTTT CTACGTTGTT GAACCACATG TACTCAGCAC	4740
	CTGGACGATT TGTCCaAGTG TTTTACATG TCACACTACA CGTATGGCAT CCTATGCATT	4800
50	TATCTAAATT TAATACCATC GCAaCTTGCG CTTTAATCTT CAAGCCAATT AACCTCCTTC	4860
	ATCTTTCTAA CTGCTACATA TAAATCCCTT TGGTTCCCAA TTGGTCCATA ATAATTAAAG	4920

55

	GGCGCGTTGT GTGAACCACC ACGTGTATCT GTAATTTCTG ACCCAGGCGT TTGAATATGT	5040
	TTATCTTG TG CATGATACAT AAACATTGTA CCTTTAGGCA TACGATGCCA AATAACTGCT	5100
5	CTTGCCGTTA CAACACCATT ACGGTTATAC ACTTCTAGCC AATCATTATC TTGGATATCG	5160
	TGTTTTTCAG CATCTTCATT TGATATCCAA ACCGTTGGAC CACCTCTAAA TAGTGTCAAC	5220
	ATATGCTTAT TATCTTGATA CATTGAGTGT ATATTCCATT TTCCATGAGG CGTTAAATAA	5280
10	CGCAGTACCA AAGCATCTGT ACCACCTTTA ATTTTCTTAT CTCTATTCCC AAATACCATT	5340
	GGCGGCAATG TCGGTTTATA TACTGGTAAG CTCTCCCCAA ATTGTTGGAA AACTTCGTGA	5400
	TCCACATAAT AACTTTGACG TCCTGTTAAT GTTCTAAAAG GTACTAGACG TTCTATATTC	5460
15	GTTGTAAATG GTGAATATCG TCGACCTTGT TTATTTGAAC CTGGGAATAC TGCTGTCCGT	5520
	ATTACTTCTC GTGGTTGTGA AGTTATATTT AAAAACGAAA TTTTCTCAGC AGCGCGTTCCG	5580
20	CTAGAAATAT CTTTTAACGG CATTCCAGTT TGTCTTCGA GATCTTCATA TGATTTTGT	5640
	GATAATTTAC CATTCTAGC AGATGAAATA CTTAGTATTG CATCAGCTAC ATTACGTGCT	5700
	GTATCAATAC GTGGACGATT CGCTCTCACA GAATCATCAT TTGTATCACT CCACGTACCT	5760
25	AACATACTTT TTAATCTTC ATATTGTTCA CTGACACCGA AACTTACACC ATGTGCTCCA	5820
	ACTTTCCCTT TTTCAAGTAC AGGACCAAGC GTGACATATT TGTCGTAAAT TTAGTGTAG	5880
	TCGCGTTCTA CAATTGCAAA GTTAGGCATT GTACGTCCAG GTACCGCTTC AATTTCACCC	5940
30	TTCGACCAAT CTTTCACTAC GCCGTATGGT GTTGAAATTT CTGCTTTGT ATCATGACTA	6000
	AGTGGAGTTG TCACAACATC TTTAAACGTT CCAGGTAAAT AGTCTTTTGC CATTCTGAA	6060
	AATGCTTTTG CCAACGTTTT ATAAATATCC CAGTCTGAAC GCGATTCCCA TAACGGATCA	6120
35	ATGGCAGGAT TGAAAGGATG TACATATGGA TGCATATCCG TTGATGATAA ATCATGTTTT	6180
	TCATACCAAG TCGCTGCCGG CAAAACAATG TCAGAATATA ACGGTGTTGC CGTCATTCTG	6240
40	AAGTCTAAAG AGACCACTAA ATCTAACTTA CCTGTTGTTT CTTACGCCA CGTAATTTCT	6300
	TCTGGCTTTT CATCTTCATT TGGTGTAGCT AATAACCCTG ATTTTGTGCC AAGTAAATGC	6360
	TTCATAAAGT ATTCTTGACC TTTGTCAGAA CTTGAAATTA AGTTTGAACG CCATATAAAT	6420
45	AATGATTTTG GATGATTCTT TTTCAAATCA GGATCTTCTA TTGCAAATTG TGTTTGTTTT	6480
	GATTTCACTT CATCAATTGC ACGTTGCAAA ATCGCTTCAT TTGAATCTAT ACCTTCATCT	6540
	TTAGCTTCTT CTGCAAACAA CAAACTATTT TTATTAAATT GTGGATATGA TGGTAACCAA	6600
50	CCAAGTCTAG CTGCTAAAC ATTATAATCA GCTGGATGTT GATGCTTTAA CTCCTCTGTT	6660
	TTAGCTAATG GAGATTTTAA ACGATCTACA TTTGACTCTT CATATTTCCA TTGGTCTGTT	6720

55

	AATGCGACAG TACTCCATCC TTCAATCGGA CGACATTTTT CTTGTCCCAC ATAGTGAGCC	6840
	CAACCGCCAC CATTACACCC TTGACAGCCA CATAACATAA CTAAGTTTAA GATTGAACGA	6900
5	TAAATCGTAT CTGAGTTAAA CCAATGGTTA ATACCCGCAC CCATGATAAT CATTGAACGC	6960
	CCTTCAGTAT CGATAGCGTT TTGCGCAAAT TCTTTCGCTA CTTGAATGAC AACACTTTGT	7020
10	TTTACGCCTG AAATGGCTTC TTGCCAAGCA GGTGTATATT TTGATTCTGC ATCGTCGTAT	7080
	CCTTTTGATT CTAATTTATG ATCAAAACGA CGCACGCCAT ATTGACTTGC CATTAAAGTCA	7140
	AAAATTGTAG CAATACGGAC TTTGTCACCA TTTGCTAAAG TGAATTGTG AGTTGGAATT	7200
15	GGACGATTGA ATATCCCATC TCCATCACTA TCAAAGTATG GGAATTGAAT TGTTTCTAAT	7260
	TCGTATCCAC CTTCTGTCAT TGATAATGTA GGGTTAATTT TAGAACCATC TTCTGTTTCT	7320
	AGTTTTAAGT TCCACTTCTT ACCTTCTTCC CAACGTTGAC CCATTGTGCC ATTAGGTACT	7380
20	ACTAACTAT CGCTGATTGC ATCATGAATA ACTGGCTTCC ATTCGCCTTG CTCTGTTGTT	7440
	TGACCTAAGT CACTCGCTCT TAAAAATCGA CCCGCTTTAT ATCCATTTTC ATCTTCATCC	7500
	AGCATGATAA GAAACGGCAT ATCTGTATAT TGTTTAGCGT AATTTATAAA GCGTTCATTA	7560
25	GGTTGATTAA CATAATGTTT TTGTAAAATA ACATGCGTCA TTGCTTGTGC AATTGCAGCA	7620
	TCTGAACCAG GATTGCGTGC TAGCCAGTTA TCTGCAAATT TCACATTTTC TCGTAATCT	7680
	GGTGCTACTG AAATGACTTT TGTACCTTTA TAGCGGACTT CAGTCATAAA ATGTGCATCC	7740
30	GGAGTACGTG TTAAAGGTAC ATTAGAGCCC CACATAATAA TGTATGATGC GTTATACCAG	7800
	TCACTTGATT CAGGCACATC TGTTTGCTCT CCCCAAATTT GTGGAGAGGC AGGTGGTAAA	7860
35	TCTGCATACC AGTCATAAAA ACTAAGCATT TCACCACCAA GCAAATTGAT GAATCGAGCA	7920
	CCTGCTGCAT AACTAATCAT TGACATCGCT GGAATAGGTG TAAATCCTGC GATTCGATCT	7980
	GGACCATATT TTTTATTGT ATACAGTAAT TGTGCTGCGA TTATCTCTGT AACGTCTTTC	8040
40	CAATTTGAAC GCACGTGCCC TCCCATACCT CGGGCTTGCT TATATTGTTT GGCTTTGTCT	8100
	TCATTTTCAA CAATAGACGC CCATGCAGCA ACGCGATTAC CATTGTTTTT TTCTAATGCT	8160
	TCAGTCCATA AATCCCAGAG TTTTCCACGA ATATATGGAT ATTTGATTCTG AAGCGGACTG	8220
45	TATTCATACC AAGAGAATGA CGCACCTCGT GGACATCCTC TCGGTTTATA TTCAGGCATA	8280
	TCCGGACCAC AACTTGATA GTCAGTTTGT TGATTTTCCC AGGTAATCAC ACCATTTTTT	8340
	ACAAATACTT TCCAAGAACA TGAGCCTGTA CAGTTAACAC CATGTGTTGT TCTTACTTCT	8400
50	TTATCGTGGC TCCAACGTTT TCTGTACATT TTTTCCCAT CTCTACTTTT ACTTTCTAGG	8460
	ATCGACCAAT TCCCATTAATA TTTTCTGTT GGCTTAAAGA AATTCAATCC AAATTTTCCC	8520

55

	TAAAATGCCC AAGACTATTG CTTTAATTAG ATTGTACATT TTTTCACAAA CATAAAATAT	8640
	TAGGGAATCA CCTAATTACT TAAGGAATTT CCCTATCAAT AACGGGATTT CATTGAAATA	8700
5	ATACACAATC ATGTATGGTC ATGCTTATTG CCAATCTAAA TCGTTCAAAT TTGGCACAAC	8760
	GACAAATAAG GCTTCAACAC GAATATATTC TCTCGTTGA AACCTTACTT ATTCATTTAT	8820
	TTTTTATAAA TTAGTGACAT AACACTGTAT TAGCATCTGC ACGATCGGTT GAAATATATG	8880
10	TTACATTTTC TTGCTGCTTA ATAAATGCAT CATAGTAATC ATATTGCGAC GAATGATATG	8940
	TGCCATTCGA TGTATCATTT GGGTTTAGCA AACAGCCATA ACCTTCGTCA TATAAATGTT	9000
	CACAGAGCAT AAGGGCGTCA TGTTTAGAAC CACTTACTAC ATAAAATTGC TTCATAGGAT	9060
15	CATATGATTT AGGAGTGTTT TCAGTATAAT CAACAACCTC CCCTATAATA CATATACCTG	9120
	GTTTCGCCTC AATTGAATAG TGTGCAATT TTGAAATAAT ATTACTTAAA CGCCCCCTTAA	9180
20	CAACAAACTC GTTAAACAC GATGCTTGAA AGACAATCGC TATCGGGTAA TCAATATCTG	9240
	TGTATTGTTG TATCTGTGTG ATAATTTTCC CTAAACGTTT TACCCCCATA TAAATTGCTA	9300
	ACGTGCCACC ATCTACTAAG GAATTGACAT CCACTTCATT TTCTTCTGAA TCTTTAAAGT	9360
25	GACCTGTAGA AAATGTCACA CTTTLAGCAA CTGTACGCAT TGTCAAACCT GTCTGCATAG	9420
	TAGCAACTGc tGCGCTCGCT GATGTCACCC CTGGTACAAT TTCAAACGCA ATATGATGTT	9480
	CATTTAGTAT GTCGACTTCT TCTTGCACAC GACCAAATAT CGCTGGATCG CCACCTTTAA	9540
30	GTCTAACAAC CTTGTTATAT CGACGCGCTG CTTCCACGAT ACAGTCAITTT ATTTTTTCTT	9600
	GCTGAATATG TTTTGCATAC GGCTTTTTAC CAACATCGAT AATTTTCAGTA GTCAAATTCG	9660
35	CATATTGTAA AATTAACGGA TTCACTAATC GATCATATAG AATGACATCC gCTTCACGTA	9720
	TTAAACGCTC AGCCTTTTTT GTCAAATAAT TCGGATTACC TGGACCCGCA CCTATCAAGT	9780
	AAACCTTGCC ATATTCCTCT ACAGACATAT ATATACGTTT CCGTCTGTAA CTTCTACCTC	9840
40	ATAAACATCT ACACAACCTT CATCAGGTTT TTGAACAATA CCTGTATTTA AATCAATTTT	9900
	TTGATCGTGG AGCGGGCAAA ATACATATTC CCCACTCACT GTCCCTTCAG ACAATGGTCC	9960
	TTGTTTGTGT GGACAGATAT TGTGAATCGC ATGAATTTTG CCACTTTCTG TTAAAAACAA	10020
45	CCCTACCTCT TTGCCTTTGA CAATAACCTT TTTTCCAATT AGGGGTGTTA ATTCATCTAT	10080
	AGTTGTCACT TTAATTTTTT CTTTGTGTTT CATGTATTAC ACCTTCTCCA CTTCAAAAAT	10140
	TCTACGTGCT TGAGCATTGC TAGTTATTGC TTCCCAAGGT TCAGCTTCGA CTGCTTTTTT	10200
50	AGCATCCATA ATGCGTTCAA ATAGTTCATT TTGTCTTTCT GGGTCAAGTA AGACTTCTTT	10260
	TACATTTTCA AATCCAAGTC TTCTTAACCA TGGCGCTGTT CTTTCAGCAT ATATACCTGT	10320

55

	AGTTGTTAAA AATTCAGCTT TTTCAACTTC TGTACCACCA TTACCACCGA TATAGATTTG	10440
	GAATCCATTT TCAACTGAGA TAATACCAAA ATCTTTAACA CCTGATTCAA CACAACCTTCT	10500
5	TGGGCAGCCT GATACACCCA TTTTGAATTT ATGAGGTGTA TCGATGTATT CAAATGTTTT	10560
	TTCTAAACGA ATGCCAAGTC GTGTCGTGTA TTGCGTACCA AATCGACAAA ACTCTTTACC	10620
	AACACAGCTT TTAACGAGC GTGTTTCTT ACCATAAGCT GATGcTGAAC GCATACCTAG	10680
10	GTCTTCCCAT ATATTTGGTA ATTCTTCTTT TTTAACTCCA TACAAACCAA CACGTTGTGA	10740
	ACCTGTCACT TTAAC TAGTG GCACATGATA TTTCTTAGCC ACTTCTCCTA GACGAATCAG	10800
	TTGGTCTGCA TCTGTAACAC CCCACGCAT TTGAGGTATA ACAGAAAATG TACCATCATT	10860
15	TTGAATATTC GCATGGTAAC GTTCGTTAGC AAATCTTGAT TCTCTTTCAT CTTCATGATC	10920
	ATGTGGATAA ACCATGTTTA AATAATAGTT GATTGCTGGT CGACATTTTG GACATCCACC	10980
20	TTTATTTTTA AAGTTTAAAA CATGTGGAAC TTCTTTAGAT GTTTTTAAAC CTTTCGCTCT	11040
	TATTTGCGTT ACTATTTGAT CGCGTGTCAA ATCAGTACAA CCACATATAC CAGCAGGTTT	11100
	TGCGGCAACA AAGTCATCTC CTAAGGTGTG CTGCAATATT TGAGCAATTT GCGGTTTACA	11160
25	TTTACCACAT GAATTCCCCG CTTTTGTTTT AGCCGTTACT TCTTCAACTG TTGTAAAGCC	11220
	ATTTTCCGTA ATCGCATTTA CTATAGTACC TTTATCAACA CCATTACAAC CACAAATTGT	11280
	TTCATCATCA GCCATATCAG CAATTGATAG CGATGCCTCT TCTCCACCTT TAGTAAGCAA	11340
30	TGATACAAGT GTGTAATCTT CAGTGGATTG ACCTTTTTTC ATCATGTTAT AAAAGCGTGA	11400
	ACCATCATCG ATATCACCAT ATAGTACTGC ACCAACTACA TTACCGTCTT TTAAAAAGAT	11460
	TTTTTTATAG TTATTATCAA CACTATTAAA TATTTCAATA CCTTTAATTT CTGCATTTTC	11520
35	TACAATTTGA CCAGCACTAT ACAAGTCACA CCCAGAACT TTTAATGACG TAAATGTTGT	11580
	TGATCCCTTG TATCCGTTTC TTTCTTTATT TGTTAAATGA TCAGCTAATA CTTTACCTTG	11640
40	TTCATATAGT GGTGCAACGA GTCCATAAAC TTTGCCGTTA TGTTCTGCAC ATTCACCAAC	11700
	TGCATATACA TTGCTATCAC TTGTTTGCAT CACATCATTG ACAACAATAC CACGATTAAC	11760
	ATCTAGACCT GATTCTTTGG CTACTTCTGT GTATGGTCGT ATACCTACTG CCATAACAAC	11820
45	TAAGTCTGCC GGAATCTCGC GTCCATCAGC CAATTTAACA CCCTCAACAT CATCTTCTCC	11880
	TAAGATTTCA GTTGTGTTGG CTTGCATTTT AAACCTTCATA CCTTGCTTTT CTAGATCTGC	11940
	TTTAAGCATA TTTCCAGCTT TACGGTCTAG TTGCATTTCC ATCAACCATT CAGCTAAATG	12000
50	TAACACCGTT ACTTCCATAC CTTGATCTAA TAAACCACGT GCACACTCTA AACCTAGTAA	12060
	TCCTCCACCA ATTACAATTG CTTTCTTTTT AGTCTTAGCA ATGTTTCATCA TTTGTTTCAGT	12120

55

	GAATGCTTTA GAACCTGTCTG CAAAAATCAA TTTATCGTAT GATACTTCAA TACCATTTGC	12240
	AGTAGTAACT GATTGATTTG CTCTATCTAC TTCAATTACA GGATCATTTG TAATTAAGTC	12300
5	GATACCATGT TCCTCATACC ACTCATATGG ATTCATAATT GTTCTTCAA CTGTCATTTT	12360
	ATTTTGTAAG ATATTTGAAA GCATGATGCG GTTATAGTTT GGATAAGGTT CTTTACCTAT	12420
	TACCGTAATA TCATATAAAT CGTTGGCGCG CTCTAATATT TCTTCGATTG TTCGAATGCC	12480
10	CGCCATACCG TTACCAATCA TTACTAGTTT TTGCTTTGCC ATAAAATATG CCCCTTTACT	12540
	CCATAATATT TATTTCAAAA AAAGGTATTA ATTTTTCGTT AGTGCTTTTA TATTTTCATT	12600
	GGAATCATTG AGCTTTCTAA TCTATCGTTA ATGATTTGCT TTAAGATTGG GTCGAAGTTA	12660
15	ATTGAAGGTG TGAAGTGTAT ATCTGTATTA ATAACCATGT CATTCAATTG CTGCTTCACT	12720
	TTGTTAACAA GTCTTCCGTC ATATAAAAAAT AATGGTACGA CAATCAATTT TTGATACCGT	12780
20	TTGAGATGCG TTTCTAAATC ATGTGTAAAA CTAATCTCTC CATATAGCGT TCTCGCATAT	12840
	GTCGGCTTGC TAATTTGCAA ATTTTGAGCG CATATTTGTA ACTCTTCGTG TGCCTTAGTA	12900
	AACTTTCCAT TAATATTGCC GTGTGCAACA ACCATAACTC CAACTTGTG TTCGTCACCT	12960
25	GCTAATGCGT CACAAATACG TTGTTCAATT AATCGTCTCA TTAAAGGATG TGTGCCAAGT	13020
	GGCTCGCTTA CTCTACCTT TATGTCTGGA TACCGTCGTT TCATTTCATG AACGATATTC	13080
	GGTATATCCT TGAGATAATG CATTGCACTA AAGATTAGCA ATGGTACAAT TTTAAATGG	13140
30	TCAACCCAC TTTGAATCaA CGTCGTCaTT ACCGTCTCTA AATCCTGATG CTCACCTTCT	13200
	AAAAACGCAA TATCATAGTG ATGTATATCA TCTTTTACTA ATTCAGAAAT AAATGCTTCT	13260
	AACGCTTGaT TCTGTCTGCC GTGCCTCATG CCATGTGCAA CAATGATATT CCCATTACCA	13320
35	TTTACCAACC CTTTCACACG TATTGTATAC CAAATCATTT TGTTTTTGTG AAAAGAATCA	13380
	CATTATAATG TAAATCAGG GAATTCCTG ATGCCTGTAG TCATGCATAT TCCTTATACA	13440
40	TTTTCCCTTT TTGTTAAATC AAAAAAGCG ACCGATATAT GAATCCCTAC TCAACATTTA	13500
	TTTGAGCAAG CATTAAATATA TCGGTCGCTT GTAGTGATA TTATTATCTT AAAATGGTGG	13560
	TTGGCCTAAT ATTGTTTCGT CAAAGCGCTC GGGTATCAAT ACTTTGCGCA TGATCACACC	13620
45	TAAATCGCCA TCATCATTTT CATGTTTCGT GTATATTTCA TAACCTCTT TTTATAAAT	13680
	TTTAAGTAAC CACGGATGCA ATCTTGCGA TGTACCTAAA GTAAGTCCG CTGACTTTAA	13740
	CGTATCTCGC AAAAATGCTT CTTCAACATA AGTAAGTAAT TGGCTACCAT AGCCTTTCCC	13800
50	TTCATACTCA GGATTTGTCG CAAACCACCA GACAAAAGGA TAACCCGAAA TACTTTTCAC	13860
	ACTTCCCCAA GGATATCTAA CCGTAATCGT AGATATAATT TCATCATCAA TTGTCATGAC	13920

55

CCAATCAATA CCTAGTTCTC TTAGAGGCGT AAATGCTTCA TGCATGAGTT CTTGCAATTT 14040
TTCTGCATCT T 14051

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1885 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

TAATCCTCAA CTnGATTAT ATGGCTTGGG CGCATATGAA CTGCTTAGTT TAGTGATGA 60
CATTCATACA GTTCGCATGA CTATCATACA ACCTCGAATA GATAACTTTT CTA CTGGAAGA 120
GTTACCAATC TCAAGATTAC TTCAATGGGG AACCGATTTT GTTAAACCCT TAGCCAGACT 180
TGCTTATAAC GGTGAAGGTG AGTTTAAAGC AGGTAGTCAT TGTAGATTCT GTAAGATAAA 240
GCATTCATGT AGAACACGTG CAGAATACAT GCAAAATGTG CCTCAAAAGC CACCACATTT 300
GTTGAGTGAT GAAGAGATTG CAGAACTTTT ATATAAACTG CCTGATATCA AAAAATGGGC 360
TGATGAAGTA GAGAAATATG CGTTAGAACA AGCGAAAGAG AATGATAAAA CGTATCCAGG 420
TTGGAAGCTA GTCACGGGAC GTTCAAGGAG AGTGATAACT GATACAAAAG CAGTCCGAGA 480
CAGGTTAGTT GAAGCGGGTT ATAAACCTGA AGATATTACA GAAACCAAGT TACTTAGCAT 540
TACGAATTTA GAAAAATTAA TCGGCAAAAA AGCATTTTCT AAAATTGCAG AAGGCTTTAT 600
AGAAAAGCCG CAAGGTAAAT TAACACTTGC TACCGAGTCT GATAAACGAC CAGCTATAAA 660
GCAATCTGCT GAAGATGATT TTGACAACT ATAAAAATTA AAAAGGACGG TATATAAACA 720
TGAAAGCAAA AGTATTAAAT AAAACTAAAG TGATTACAGG AAAAGTAAGA GCATCATATG 780
CACaTATTTT TGaACCTCAC AGTATGCAAG AAGGGCAAGA AGCAAAGTAT TCAATCAGTT 840
TAATCATTCC TaAATCAGAT ACAAGTACGA TAAAGCCAT TGAACAAGCT ATAGAAGCTG 900
CTAAAGAAGA AGGAAAAGTT AGTAAGTTTG GAGGCAAAGT TCCTGCAAAT CTGAACTTC 960
CATTACGTGA TGGAGATACT GAAAGAGAAG ATGATGTGAA TTATCAAGAC GCTTATTTTA 1020
TTAACGCATC AAGCAAACAA GCACCTGGTA TTATTGACCA AAACAAAATT AGATTAACGG 1080
ATTCTGGAAC TATTGTAAGT GGTGACTATA TTAGAGCTTC AATCAATTTA TTTCCATTCA 1140
ACACAAATGG TAATAAGGGT ATCGCAGTTG GATTGAACAA CATTCAACTT GTAGAAAAAG 1200
GCGAACCTCT TGGCGGTGCA AGTGCAGCAG AAGATGATTT TGATGAATTA GACACTGATG 1260

TTGAGGTGTC AAGAATTTGA AATTTATGAA TATAGATATT GAAACATACA GCAGTAACGA 1380
 TATTTTCGAAA TGTGGTGCCT ATAAATACAC AGAAGCTGAA GATTTTCGAAA TTTTAATTAT 1440
 5 AGCTTATTCG ATAGATGGTG GAGCGATTAG TGCGATTGAC ATGACTAAAG TAGATAATGA 1500
 GCCTTTCCAC GCTGATTATG AGACGTTTAA AATTGCTCTA TTTGACCCTG CTGTAAAAAA 1560
 GTATGCATTC AATGCTAATT TCGAAAGAAC TTGTCTTGCT AAACATTTTA ATAAACAGAT 1620
 10 GCCACCTGAA GAATGGATTT GCACAATGGT TAATTCAATG CGTATTGGCT TACCTGCTTC 1680
 GCTTGATAAA GTTGGAGAAG TTTTAAGACT ACAAAGCCAA AAAGATAAAG CAGGTAAAAA 1740
 TTTAATTCGT TATTTCTCTA TACCTTGTA ACCAACAAAA GTTAATGGAG GAAGAACTAG 1800
 15 AAACCTACCT GAACATGATC TTGAAAAATG GCAACAATTT ATAGATTaCT GTATTCGAGA 1860
 TGTAGAAGTA GAAATGGCGA TTGCT 1885

20 (2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2656 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

30 TAATCCTTAG TTCACTGnCA AATTTCAAAA CACCAGTTC CTCTATCTGC ATCCATAGAA 60
 ACTGnATGTT TGTGTCAATA ACCGATTAT ATTGTGATGn TGTTTGTAAC TCGATTAAGT 120
 TATCATCTTT CGAAAAATTA TCTACTACCA TTATTCAACC ACCTTTCCTT CGAATAAACT 180
 35 CCATTTACCA ACKCCACCAG TACCAAAGTT TCTAACTAAA AATTGATGTG CAGACGGGAA 240
 GTTATTACGT CTTAATACTT GTGTTGTATT ACCTGGTGTA TTCGATTTTA CTTCTAATAT 300
 CCAACCTGCA ATACCTTTAA AGTCTTTAGG AAAATCAGTA AATCGGTTTG ATTCTTCAGT 360
 40 AGTGATATAG AAATCTAAAC CAACGATTTT TAAATCTGAT AATTTTGTAA TACTCTTAGG 420
 GATATGTTCC CAATAACCGG CGTTTTGCGG GCAGAAATTC CATGCTCCGT TGTTTTCTT 480
 45 AITGAAAATG TCAATGACAC GTTCGAATTT AAGCATATTT CTACCTGTGC TGTTTCTGGT 540
 AAGTACTTGT CTTAGAGCAC CATTATAGTG TCCAGGCAGT ACATCCAAGA ACCACCCTGC 600
 ATCTCTAAAC GCTTTCGGTA ACGGGAAATC TAATGCATTT TGTGTGTCTT GaCGTATAGA 660
 50 TATAGTAATG ACCAACTTCC GTAATATCAC TTAGATATGC TGGGTTCTGT ATTGGTAACG 720
 GTTTAACACG TCCGCCTGAA TCAGTCATTG ATACTTGAGG TGCGATGTTT TTCAAGAATT 780

55

	TAGTTACCCC GATTAGAAGT GCTTTACGTC CTGTTTCTAG ATCGTAATAC ATATCTAGAC	900
	CCTCAGCCTC TTGGAAATCT CCTTTAAAGT TGTTATTTCAC ACCGCCTATA TCGATGCGAC	960
5	GTTTAAATAA CAATTCTTTC GTTTGTATAT CGAAGCCTTG TAAGTAGTTA GGGTTGGCTG	1020
	TATTCGAATC ACCTGTATAC CAATATAAGA TACCTGCATC ATAAGTGATA CCTTGCATAG	1080
	GTTGTGTATC TGAAGTGTAT TCCATAGGTA TATCCATTG ATACAATACT TTGTCTATAC	1140
10	CTTTATCAAT ATCGTCAGCA CTTCTAACCT CAACAAAGTT CAACGAATTC TTAAGTTGTC	1200
	TTTCAGTGGG TTTATATTCA CGTCTAAAA TCATTAAATT TTCTACCGGA TTATAAATCG	1260
	CTGACGTATA TCTGTGCTTA AATATATTTCG GCATGACATC TTGCATTTC TTACCATAAG	1320
15	TTATTTCTCC AGTTCTATAT TGGAAACGTA CAAACTTGTT GTTTTTGTGA CTGTCCAATA	1380
	CAGCTGAATA AATCCATAAT TCTCCATCAA TGTATCTATA CGCATTGTGT GTACCGTGAC	1440
20	CGCCGTTTTT AACAAGCAAT CTATCAATAA ATTGTCCGTT GGGCTTCAAT CTAGATAACA	1500
	TGTAATGATT ACCTGGACGA GCTTGCCTCA TATAAATAAT TTTCGTTCTA GGGTCTACCC	1560
	AAAATGATTG CATTACTGCA TTTGTATATG GCGATAAATC AGTGATAAAT TCCGGTTCTT	1620
25	GCTCTTTTGG TTCGAATCGG TATTCTGTCT CTCGATATTC TTTATAGTGT TCATCTACAG	1680
	CTTTCTCAAC CTTTTTAGTG AAAACATCTA GTGTTGAATA ATCATGATAC AAACGATCTT	1740
	GCAATGTCTT ATGACCATAA CCTGTATTAT CAACGCGCGC GTCTTTTACT TCGTTGATAC	1800
30	CGTCGCCGTT ATGACCTAGT ACCATGTTGC TAAATCGACC GTTTAAATAT GTTAAAAAGT	1860
	CAGAGACGTT ACTTGTAACA TTTAAATGTT CATACTTTAT TTGTTCTCCA TCATGTGCGA	1920
	ATACCTCTTT ATTTCTGTGG TATTCAAGAG AGAAATTAAA ATCCGTCAGC ATGTCTGAAA	1980
35	TAAGTTTAAA GTTATACTCA TTTTCATCTA CATATCTGTA GTCAAAGACT CTACTTAAAT	2040
	CTGTAAATTAG TTTATTACTC ATGTTTTCCT CCTTTACTAT CCATAAAACT GATmATAATT	2100
40	TTTAATAAGC TCATACATAA TAACTTCATG ACCTCTTTC TTAGGATGTA ATCCATCAGG	2160
	CATGCTAGAT TTTCTAAATG CTGGATTATA TGGTTTGAAA TAATCTGTGT GATAAGCATC	2220
	ATATACTGGT ACATCCAATT CACTACAAGC CAATATCTGA GCATTGACAT AATCCTCTAA	2280
45	AGTTAACCCT AGTTTGTTTT TGTCCGTATC TTTACGGCGT ATCGTTGTAC CACTCATAGG	2340
	GCATTGCCTA GTAGCTGTCA TTACAAGTAT TTTTGAAGCT GGATTATTTT TCCTGATAAC	2400
	TTCAATTGCA GAACAAAAGG CGCCGTAAAA CGTTTTAGTG TCGGTTTTAT CAGTGCCTAT	2460
50	CGGTACGCCT GCCCAATAAC CATGTAACCA GTCATCATCT GTACCTTGTA ATATGATTAG	2520
	GTCTCCTCTT ATTTGCTCTG CTTGTCTaTA AATGCTGTTT TCTaCCGCTT CTTTACCTAT	2580

55

CTTGCCTAAC ATTTCT

2656

(2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4854 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15	AAAATGAGGG TTCTAGCGGA AATTACCAAA AGCGTGGTTC ATACTATGGG CAGCGTAATC	60
	GTATTTTCAA AGAAAAACA CCTAAATGGT TAGaAAATAG AGATAAACCT AGTGAAGAAG	120
	ATTCGGCTAA AGATAATAGC GTAGATGATC AACAATTAGA GCAAGATCGA CAAGCATTTTC	180
20	TAGATAAATT ATCTAAAAA TGGGAGGAGG ACAGTCAATA ATGAAGCAAT TTAAAAGTAT	240
	AATTAACACG TCGCAGGACT TTGAAAAAG AATAGAAAAG ATAAAnCAGA AGTAATCAAT	300
	GACCCAGATG TTAAGCAATT TTTGGAAGCG CATCGAGCTG AATTmACGAA TGCTATGATT	360
25	GATGAAGACT TAAATGTGTT ACAAGAGTAT AAAGATCAAC AAAAACATTA TGACGGTCAT	420
	AAATTTGCTG ATTGTCCAAA TTTCGTAAAG GGGCATGTGC CTGAGTTATA TGTTGATAAT	480
	AACCGAATTA AAATACGCTA TTTACAATGC CCATGTAAAA TCAAGTACGA CGAAGAACGC	540
30	TTTGAAGCTG AGCTAATTAC ATCTCATCAT ATGCAACGAG ATACTTTAAA TGCCAAATTG	600
	AAAGATATTT ATATGAATCA TCGAGACCGT CTTGATGTAG CTATGGCAGC AGATGATATT	660
	TGTACAGCAA TAACTAATGG GGAACAAGTG AAAGGCCTTT ACCTTTATGG TCCATTTGGG	720
35	ACAGGTAAAT CTTTTATTCT AGGTGCAATT GCGAATCAGC TCAAATCTAA GAAGGTACGT	780
	TCGAaAATTA TTTATTTACC GGAATTTATT AGAACATTAA AAGGTGGCTT TAAAGATGGT	840
40	TCTTTTGAAA AGAAATTACA TCGCGTAAGA GAAGCAAACA TTTAATGCT TGATGATATT	900
	GGGGCTGAAG AAGTGA CTCC ATGGGTGAGA GATGAGGTAA TTGGACCTTT GCTACATTAT	960
	CGAATGGTTC ATGAATTACC AACATTCTTT AGTTCTAATT TTGACTATAG TGAATTGGAA	1020
45	CATCATTTAG CGATGACTCG TGATGGTGAA GAGAAGACTA AAGCAGCACG TATTATTGAA	1080
	CGTGTCAAAT CTTTGTCAAC ACCATACTTT TTATCAGGAG AAAATTTTCAG AAACAATTGA	1140
	ATTTTAAAT GATTGGTGTA TAATGAATAC AAATCTAAAT CGTTTAAATG ATTGAAGACA	1200
50	AGATGATCTA ATCAATATTA CACAGAAAGC CATTGTTTGA TGAGAATATG GTTAATAAAT	1260
	TAGATGATTA CTACTTCATT TATGGTATTT GTAATGAATA CCCGGATCAA GACCGTTATC	1320

	CTCGTCCCTT	GTATAGGGGC	GGGATTTTTT	GTTTTTTTCA	GACATAAATG	TTTGTGGTG	1440
	TCATAAATTC	CCTGTTTATT	GTTAATAGGT	TTAATGTTAA	AACGATGATT	GTTGTTCAAT	1500
5	TTTTTAACGA	GGTCAGATAA	AAGTATTTAT	AAAGCAAATA	GGAGGGTTTA	ACATGGAACA	1560
	AATTAATATT	CAATTTCCAG	ATGGTAATAA	AAAGGCGTTT	GATAAAGGTA	CTACTACTGA	1620
	AGATATAGCA	CAATCAATTA	GTCCTGGATT	ACGTAAAAAA	GCTGTTGCCG	GCAAATTTAA	1680
10	CGGGCAACTT	GTAGATTTAA	CTAAACCGCT	TGAAACTGAT	GGATCAATTG	AAATTGTGAC	1740
	ACCAGGTAGT	GAAGAagcGT	TAGAGGTATT	ACGTCATTCT	ACTGCACATT	TAATGGCACA	1800
	CGCGATTAAA	AGGTTATATG	GTAATGTTAA	ATTTGGTGTA	GGTCCTGTAA	TAGAAGGTGG	1860
15	ATTCTACTAT	GACTTCGACA	TTGACCAAAA	CATCTCATCT	GATGACTTTG	AACAAATTGA	1920
	AAAAACAATG	AAACAAATCG	TTAACGAAAA	TATGAAAATC	GAACGAAAAG	TGGTTTCACG	1980
20	AGATGAAGTG	AAAGAGTTAT	TCAGCAATGA	TGAATACAAA	TTAGAATTAA	TCGACGCGAT	2040
	TCCTGAAGAT	GAAAATGTAA	CATTATATAG	TCAAGGTGAT	TTTACTGATT	TATGTCGTGG	2100
	AGTTCACGTT	CCATCAACAG	CTAAAATTAA	AGAGTTTAAA	CTATTATCTA	CAGCAGGTGC	2160
25	ATACTGGCGT	GGAGATAGTA	ACAACAAAAT	GTTACAACGT	ATATACGGTA	CTGCTTTCTT	2220
	TGATAAAAAA	GAATTGAAAG	CACATTTACA	AATGTTAGAA	GAGCGTAAAG	AACGTGATCA	2280
	TCGTAAAATT	GGTAAAGAGT	TAGAACTATT	CACAAATAGC	CAATTAGTTG	GTGCTGTTTT	2340
30	GCCATTATGG	TTACCTAACG	GTGCAACAAT	TAGACGTGAA	ATTGAACGTT	ACATTGTTGA	2400
	TAAAGAAGTT	AGCATGGGAT	ATGACCACGT	TTATACACCA	GTACTTGCTA	ATGTTGATTT	2460
	ATACAAAACA	TCTGGTCACT	GGGATCACTA	TCAAGAAGAT	ATGTTCCAC	CAATGCAGTT	2520
35	AGATGAAACT	GAATCTATGG	TATTACGTCC	AATGAACTGT	CCACATCATA	TGATGATTTA	2580
	TGCGAATAAA	CCACATTCAT	ATCGTGAAAT	ACCTATCCGT	ATCGCTGAGC	TAGGAACGAT	2640
	GCATAGATAT	GAAGCAAGTG	GTGCTGTATC	AGGATTACAA	CGTGTTCGTG	GTATGACTTT	2700
40	AAATGATTCA	CATATCTTTG	TTCGACCTGA	TCAAATTAAA	GAAGAATTCA	AACGCGTTGT	2760
	AAACATGATT	ATTGATGTGT	ATAAAGACTT	TGGTTTCGAG	GATTATAGCT	TTAGATTAAG	2820
45	TTATAGAGAC	CCTGAAGATA	AAGAAAAGTA	CTTTGATGAT	GATGATATGT	GGAATAAAGC	2880
	TGAAAATATG	CTTAAAGAGG	CAGCGGATGA	GCTTGGCTTA	TCGTACGAnG	AAgCGATTGG	2940
	TGAAgCGGCA	TTCTATGGTC	CGAAACTAGA	TGTTCAAGTT	AAAACAGCGA	TGGGTAAAGA	3000
50	AGAGACATTA	TCAACAGCAC	AACTTGATTT	CTTATTACCA	GAACGTTTTG	ATTTAACTTA	3060
	TATTGGTCAA	GATGGTGAAC	ATCATCGTCC	AGTTGTTATT	CATCGTGGTG	TTGTATCAAC	3120
55							

	AGCGCCAAAA CAAGTTCAAA TCATTCCAGT TAACGTTGAT TTACATTATG ATTATGCGCG	3240
	CCAATTACAA GATGAATTGA AATCTCAAGG CGTTCGTGTA AGTATTGATG ACCGTAATGA	3300
5	AAAAATGGGT TATAAAATCA GAGAAGCTCA AATGCAAAAA ATACCTTATC AAATCGTAGT	3360
	TGGGGATAAG GAAGTTGAAA ATAATCAAGT GAATGTGCGT CAATATGGAT CGCAAGACCA	3420
	AGAAACAGTT GAAAAAGATG AATTTATCTG GAATCTAGTT GATGAAATTC GTTTGAAAAA	3480
10	ACATAGATAG ACAGTTGTCG CAATAAAATG CTTTAAAACT TTTATTGCGT ATCAAGTTTT	3540
	ACAGGGTTGA TTATGCGTGA TGAATCCTGT ATATTACAAG TTAGTTAAAA TATTAAATTG	3600
	AGTTAGAGGT TGCATGTTTA ATTAGTAACT TGTCAGAAGT ATTTATGGTA CATAAGTTGA	3660
15	ACAAGTGAAA GGTAAAGATG CCGAAATAGA TATAAACCAT AAATTATATC TATTGGGACA	3720
	GTTTTCGAAT AGGAACTGTA CTGTCACAGA ATGTGATGTG CTACCTTATA TAGATAATTG	3780
20	CCAAAGTGGT TGCATATCTT AAAGGTATGT AGCCACTTTT TTACTTTTAA TATCACTATG	3840
	TTCTGTAAAA AAGGGTATGA AAGTGAATAA AGGTTATTTA TTTCTTGGCC TCTAAAACAT	3900
	GGAAAGGGAG CTTATATGTC AAAAGTTCAA AATGAAAGTA ACAATGTTGT CAAAAGGGGA	3960
25	CTTAAAGATC GTCATATTTT TATGATTGCG ATTGGGGGTT GTATTGGTAC AGGTTTATTT	4020
	GTAACCTCTG GTGGAGCAAT TCATGATGCA GGTGCTTTGG GTGCATTAAT AGGATACGCA	4080
	ATTATCGGAA TAATGGTATT TTTCTTAATG ACGTCACTTG GCGAAATGGC TACGTATTTG	4140
30	CCAGTATCAG GTTCATTTAG TACATATGCT ACAAGATTG TTAGTCCATC TTTAGGGTTT	4200
	GCGCTTGTTT GGAACATTG GTTTAACTGG GTAGTGACTG TAGCAGCAGA TATTACGATT	4260
	GCAGCACAAG TCATTCAATA TTGGACACCA TTGCAAGGCA TACCCGCTTG GGCATGGAGT	4320
35	GCGTTGTTCT TAGTTATAAT TTTTAGTCTG AATTGCTTAT CAGTTCGCGT CTATGGTGAA	4380
	AGTGAATACT GGTGGCATT GATAAAAGTG GTTACAGTTA TTGTTTTCAT TGCAATTGGT	4440
40	TTATTAACGA TTGTCGGAAT CATGGGTGGT CATGTTGTAG GATTCGAAAT ATTTAATAAA	4500
	GGTGAAGGTC CAATTCCTGG TGGCAACTTA GGAGGAAGTT TGTATCAAT TCTAGGTGTA	4560
	TTCTTAATCG CTGGTTTCTC ATTCCAAGGT ACTGAGTTAA TTGGTATTAC GGCTGGTGAA	4620
45	TCAGAAAATC CTGAACGTGC TGTGCCGAAA GCAATTAAAC AAGTATTCTG GAGAATTTTA	4680
	TTATTTTACA TTTTAGCCAT TTTTGTATC GGTATGTTAA TTCCTTATGA TAGTAGTGCA	4740
	TTAATGGGGG GTAGTGATAA TGTAGCAACG TCTCCATTCA CATTAGTGTT TAAAAATGCT	4800
50	GGATTTGCGT TTGCAGCATC ATTTATGAAT GCAGTCATTT TAACGTCTGT GTTA	4854

(2) INFORMATION FOR SEQ ID NO: 107:

(A) LENGTH: 2488 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

10	ATCAAAAATT GATTGTTTTTC nATTMTTGTG TTCAGCGCGG GATCTTTTAC GTCTTTTGTG	60
	AAAACGaTTT TATTATTAAC TACTTTTACT GGATAACTTT TGTATGTCGA GTCAGTAGCA	120
	TTTTTTCTAT CGTTTGTAGT TGTGTCATAT TCACCAgTTA TTTTATGTGT GTTCTTATCT	180
15	ACCTTTAACA ACATACGGTC TTCTTTTAAA AGCTCATCTG ATCCAACAAC TGAATAAGAG	240
	GATTCTATAT ACCATGTGTC TTGATCATT TTTTCATAAT GGGGATTATC GTGACCATCA	300
	ATTTCATAAA GCGTTTCTAA GTTTTTAATA GGATACGTAC TTAGTACTTT TTTAAGACCA	360
20	TCTTTCAAAT GAATTGTTC CCACTTCATT GCCAAAAACA TATCGCCACT GACTACAATT	420
	GAAATAATAA TAATTGCTGC TAAGTTTAAAC CAGAAAATTT TATGTGCTTT CATACTTCC	480
	CACCGTTTCT CAAAATACTT CATTAACACT ATAATAATAT ATTTTGAAAA ATATTTACAT	540
25	CAGTATTAAA GTGAATATCA AATTTTAAAT TTATGAAAAT AATAGATATT TATAAAAAGC	600
	GGAAAAGAGA TACAATAAAA AACTGCATGA CGTTTGAGAC GTCACACAGT GTAACATAAA	660
30	ATTTAAAAAG TTGTTGCTAA TTTTTCAGCA TTATTAATAC TAGTTGCTTT AATTTCTTCA	720
	GTCTTATGAG GTTCAGCATT GTGTCCTTCA ATAATGATTG TTTCATATGA TGGCACACCT	780
	AAGAATGTCA TAATTGTTCT TAAATAACGG TCACCCATTT CAAAATCAGC AGCAGGTCCT	840
35	TCAGTATAAT ATCCACCACG TGATTGAATG TGTAATACTT TTTTGTCACT TAGTAAACCT	900
	TGTGGTCCTT CAGCAGAATA TTTAAAAGTT TTACCTGCAA TTGAAATAGC ATCAATATAT	960
	GCTTTAACTA CAGGTGGGAA AGAAAGGTTT CACATAGGCG TTACAAATAC ATATTTATCT	1020
40	GCACTTAAAA ATTCTTCTAA AATGTCACCT AATCTTGAAA CTTTCATTG TTCATCATCA	1080
	GTTAACGTTT CGCCATTACT CATTTTTCCC CAACCAGTTA ATACATCTTT GTCAATAACT	1140
	GGAATATAAG TTTCaATAA ATCAATATGT TTCACTTCAT CATCAGGATG TTGTTGTTGA	1200
45	TATGTTTCGA TAAATGCTTT ACCAGCCGCC ATAGAATTTG ATACCAGTTC ATTAAAAGGG	1260
	TGTGCTGTAA TATATAATAC TTTTGCCATT TGAAAAATTCT CCTCTGkTTC TGTTATTTTC	1320
	TTAAGTATAA TTATTATACT CGATATAAAA TTTAATATCA ATCAAAATAT TCAAATTACC	1380
50	ATCATTTTCT TCATCTATAT nTGGCAGTAC TACTAAAGTA TGAGTGCATT TAATTATGAa	1440
	ATAGTTGATT TaGAATATaT ACTTAATACC CAAAATATAT GAAGGATGGA TGCCACTATG	1500

55

ATTATTTATA TAGATGACAT TCAAAAATGG TTTAACCAAT ATACCGATAA ATTGACACAA 1620
 AATCATAAAG GACAAGGACA CTCAAAATGG GAAGACTTTT TTAGAGGGAG TCGGATTACT 1680
 5 GAGACTTTTG GTAAATATCA ACATTACCA TTTGATGGTA AGCATTATGG CATTGATTTT 1740
 GCATTGCCAA AAGGTACACC AATTAAAGCG CCGACGAATG GTAAAGTAAC ACGTATCTTT 1800
 AATAATGAAT TGGGCGGCAA GGTATTACAG ATTGCCGAAG ACAATGGAGA ATATCACCAG 1860
 10 TGGTATCTAC ACTTAGACAA ATATAATGTC AAAGTAGGTG ATCGAGTCAA AGCAGGTGAT 1920
 ATTATTGCAT ATTCAGGCAA TACAGGTATA CAAACGACAG GCGCACATTT ACATTTTCAA 1980
 AGAATGAAGG GTGGCGTAGG TAATGCATAT GCAGAAGATC CAAAACCGTT TATCGATCAG 2040
 TTACCTGATG GGAACGTAG CCTATATGAT TTGTAGTTAT AGAAGGGTGC CCGCAGTCTA 2100
 AAAAATTAAG CAATCATTGT GTGAGTATGA TACTTACATA ATGGTTGCTT TTTTCAATGA 2160
 20 AAATCGTAAT GCTAAGTCAT ACTTGTTTGA TTTAGATATT ACTTAAAATG TAAGACAAGG 2220
 TTGTTAGCAT TGGCAGTGAA ATATCGCACA TAAAAACAT TATTGTCACA CTAGAAAATA 2280
 GTTGTGCACT ATATCAATTT TCTGTATAAA AGTTTAATTC TGACAGTAAT GTAAACGTTT 2340
 25 ACAATTTATG ATTGACATTA ATAATGACTG AATATATGAT TTATGTAAGT ATTTGTGCAA 2400
 CGTTTTTACA AAGTGTATTG CACaYCAAA CTGtAAACaA aGTATGGGGg GCCATAACAT 2460
 GGCAGAACTA AGTTAGAGCn TATTAAAA 2488
 30

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 4093 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

TTTTCTTTAT TTCAAmCTGT ATATTaATGA TGTCACTTCA TTTGATACGA TTCTTGATAA 60
 45 CCTATTCAAA ATTCCGCCAA ATAACATAAA TATTATATAA ATGCCGATAC TTTTAATCAT 120
 TTTCTACTTT TTCTTCGATA CGGAAACTTG TTTTCGAATT GAACACTTCA CCAGCTTTTA 180
 AAATTGACGG TGCTTTTTTCA CCATATAAAT TAATATCATT TGGTAAAAAT TGTGTTTCTA 240
 50 ATGTAAAGCC AGAATGTGGT TTATAAATAT TAAATGGAAT ATCCCACTCA TCAGGCTGGT 300
 TAAAAGTAAA GAACACAACA TGAGGCATAT CTGTATCGAC CTCTAACATA AATTCATGAT 360
 TTTCAACATA CATTTTATGT TCACCAACTG TAAATGGGTG ATCGAGACCA CCAAACGTG 420
 55

	TATCTTCAAA CACTTCATGT AAATCTAGAA TATCACCTGT AACAAATATTT CGCTCATCTA	540
	ATACATACAT ATCTAATTGA TTAAGTTGAAA TGCGATGATT ATCAACGACA TTATTATCTC	600
5	GATTCAAATT GAAGTACACA TGATTGCTAG GACTAAACAA TGTGTCTTCT GATGCAACTG	660
	CTTCGTATTC AATCGACCAT TGGTGATCCG CATCATAAAT ATGTGTAATC GTCACATCGA	720
10	TATCACCCGG GAAATGATCA TCAGCTGATT TCAACACCGT CTTAAATATA ACTTTAATTT	780
	GAGCAATTTT ATTTCTAATT TCATAATCAA ATAACCTTATT GTCCAAACCA TGACATCCAC	840
	CATGTAAATG ATGTTACCG TTGTTTTTTT CTAAGTATA TTCTTTACCT TTCAACTTAA	900
15	ATTTAGCATT ATCAATTCTA CCGCTATATC TTCCTATAGA AGCACCAAAT TTAAAAGGAT	960
	TACTATGATA AAATTCATCC GCTTCAACAA CATTTCCAAG AACAAATATTA TTATCATGAT	1020
	ATTTCCAAGA CACTACTCTT GCTCCATAAT TCGTAAAAAT AATTTTAGTT TCATCATTAT	1080
20	CAATTTTGAT TAAATCTACA CCTTGTCTTT GGTGCTCAAC TTCAACTATC ATTTTACTT	1140
	CTCCCTTCTA ACCACAAGTG TTCAAGCTCT GCTGGGTAGC AACATTACTA AAACACCTAC	1200
	AATACAAATG ATTGCACCGA TAACATCATA TTTATCTGGC ATTTGTTTAT CTACGACCAT	1260
25	CGCAAAAATC AAAGTCATGA TGATAAATAC GCCACCATAT GCTGCATATA CTCTTCCGAA	1320
	TGATGGAAAT GATTGAAATG TCGCAATGAC ACCATATAAC ATGAGTATCG CACCGCCTAT	1380
30	TAGCCCAACA AGTGAAGACT GTCCTTCCCT AAGCCACAGC CAAATCAGGT ATCCCCCACC	1440
	TATTTTACAT AAGCCAGCTA ATATAAATAT AAAAATCGGA TATAACATGA AATCACTCCA	1500
	TCACACATTT GCTATCAATA ATCTATCGGC TACATATCAT TTGTTTACAT TTCTTCTTAC	1560
35	TTACATTTCC CATTTTAAAA AGTTCGTTTT CACATTCATA TTGTACACTT TTTTAGACAT	1620
	TATTCTATAG CTAAATATAA AAAAATAAGA GTAACACGCT TTCATCATCA TTTTATATGA	1680
	TAAATGTGTG TCACTCTCAT CAATTTTATT TTTTAAATAC ACGTTTCATT GAATTAAATA	1740
40	AGCCACGTTT AAATGTAAGT ACTGAATCTT TATATGTTTT AATTGCAATC CATATCAAGA	1800
	CAGCTACCAT TACAATTGAG ATTAAAGAAC TTAAGATGAC CTCATATATT TGAAGCCCTG	1860
45	AAGTTTGAGC GCGTACAAC TATTGAAATG GCGCTAAAA CGGAATATAA CTTGTGATTA	1920
	AAGCAAGTTG TCCATCAGGA TTATTTATCG TGAATATCGC GATATAAAAT GCAATCATAC	1980
	CAAGTAATGT CAGTGGCATC AAAGATTGAT TTAAATCTTC TATTCTAGAT GTTAATGATC	2040
50	CGAGGATGGC TGCAAGTAAT ACATACGCCG TAATTCACAC AATACTACTT ATAATCCGA	2100
	CAATAATAAT TTGCCAAGAC AATTGATTCA TTTCCACGTT AAAACCTTGT AGCAAGTCTT	2160
55	TTAAGTCAAA GGCAAAAATG CATATAACTG CCATCAATAC AATTAAAATA ATCTGAGTCA	2220

	TAATAATCAT TTCAATGACA CGCGATGTTT TCTCACTAGC AATTTCCATA GCTATTTGAG	2340
	ATGCATAATT TAAAACAATG AAGAACATTA GAAAGATAAT GCCATmaGcT AAAGCATAGT	2400
5	TGAAAATCTT TTGTCCTTCT GATACTTTAT CGACTTCATC ATTAGAAATC ACCTTATTAT	2460
	CAACTTTACT TTGTGCTTGT AATTTTTGTA AGTCTTCTTT GTTGATATTT AATTCCCCGG	2520
	CTACCATATT TGTTTGAATA GCTGTAAGCA GTGCTTGAC TTTTGTGAA TCTTCATGAC	2580
10	TTACTCGCTT CTCACTAATG ATTGTCCCTT GTAACGTGCG ATTTTGATTC ACCTTGATAA	2640
	TATAAGCTTT ATCAAGTTTA TGTTTTTTTA CTTCTTTTTT AGCATCTTCT ATAGAACTT	2700
15	TAGTAACTT AGCATCACTA TGAAATGTAT TCGCCTGTTG CTTGAAAACC TTATAGATTT	2760
	GTTCAATCGG TGCTGCTACA CCAATTTTAT CTGGACCATC ATCAAACATG TTAATAATCT	2820
	TATCAATGTT AGATAGGCCA ATCATTAAAG CAGCAATAAT AATCATAAAA ATTACAAATG	2880
20	ATTTAGCTTT AATTTTTTTG ATATATGTCA AAGTAAATGT CGCCCCAAAAC TTATGCATCC	2940
	TTGCCACCAA CCTTCTCAAT GAATATATCT TGTAATGATG GTTCTACAAC TTGGAATCGT	3000
	TTAACATAAC CTTGATGTGC CACAACTTGA TAAATATCTT TGGCTACGTC TTCATTCTCA	3060
25	ATCGTCAACT GAAGACCTTG CTTCATGTTT TCACTATGAA TGATGCCTCT AATGTTTGTT	3120
	AAATCTGGTA GTGTTGTTT TGATTCAATG ACAACTTTCT TGTTACCATT AGATGCACGT	3180
30	ACATGATTGA TATCACCAGA AACAACAAGT TGACCTTTAT CTAAAATACA AACATCATCA	3240
	CATAATTCTT CAACATGCTC CATACGGTGA GAACTATAAA CGATTGTACT GCCCCAATCA	3300
	TTTAAGTCTT TAACTGCTTc TTTTAATAAC TCAACATTAA CTGGGTCTAG ACCACTGAAA	3360
35	GGCTCATCTA ATATTAGTAA TTCTGGTTTA TGTAACATAC TTGCTAACAG CTGAATTTTT	3420
	TGTTGATTCC CTTTTGATAG ACTATCAATT CGTTTTTTGC GGTTTTCAGT AATATCAAAA	3480
	CGCTCAAGCC AATACGATAT TTGCTGTTGT ATTTCTGTTT TTGACATTCC CTTTAAAGTT	3540
40	GCCAAATATT TCAATTCTTC TTCAACTGTC AATTTCCCAT GTAAACCGCG TTCTTCCGGT	3600
	AAATAACCAA TACGATTGTA CATTGTTTTA TCTAGTTTTT TACCGTTATA CGTrrTGTGT	3660
	CCTTCAGTTG GTTCACTTAA GCCTAAAATC ATACGAAATG TCGTTGTTTT ACmTGCACCA	3720
45	TTTCTCCTA GAAAACCTAA CATTTTACCT GATTCTAACT TTAATGAAAT ATCATTTACT	3780
	GCCGTCATCT TGCCAAAACG TTTGTAACA TGTTCAATTA CAAGTCCCAT ACTTTGCCTC	3840
50	CTAAAAAnAT ATGTATTTAT CTTAATATAA CATTTCCATT CTCTATAAAT GCAATATTTT	3900
	TAAATGAAT TTATTTTTAA AATTTCTGAA ATTGAAAAAT TTAAATAGTG CCATTTTTGC	3960
	ATGTTAAGTA TCATTAGCAC TAGATATGTT TTTTCCATGC CTTTATTGCC TTATTTGTAA	4020

55

CTTnCCGGTG TTT

4093

(2) INFORMATION FOR SEQ ID NO: 109:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17846 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

15	TGCCAAACTA CCTTTTGACA GTCGTTGCTG TACTTCAGGA TGATCAATCA CATATnTTAC	60
	TTTATCAAAT AGGGCATCTT CATCATTTTT AGTAATTAAA TAACCATTGA AATCTGAAGT	120
	AATCAGTTCG TTAGGTCCAT ATTTAATATC ATAACATAA ACTGGAACAC CATGTGCTAA	180
20	AGATTCAAGT AGCGCTAAAG AGAAACCTTC CATGTTACTT GTTATTAAAC TCAAATAGGC	240
	ATCGCTATAT TCTTGGTCTA GATTGCTTAA AAAGCCGCGT AAGTAAACAT GATTTTCCAA	300
	TCCATATTTT TGTATCAATT CATTTAATTT TTTACTTTCA GAaCCAAAAC CATACATATG	360
25	AaGCTCTATT TTTGGGACAT ACGATACTAA GCGTTTAATT AATTCAATTT GTTGATGTAA	420
	TTGTTTTTCA GGTGAATAAC GAGCAACGGA AATTAATTTA ACACTGCGCT GATCTAATGT	480
30	TTGGACTGGT GTATCAATTG TTTCACTATA GCCGACAGGA ATATTAACAA CTGGAATAGT	540
	ATGGTTAATA CGTTTTTCAA CATCTAATTT TTGCTGCTCA GTAGAAACGA TAATTGCACG	600
	ATATCGAGAT AAATTTTCAA ACATCGCTTT ATATACATTT TTAAATGGCG ATGAATCTAA	660
35	TGCATCAATA TTTTAAATGT GTGTACTGTG AAGCACAGCT ACTACTGGGA TTGACTCAGG	720
	CGTTAAGTTG AAAATAGGTG CTGTGTACAC ATTACGATCA CTGAAAAATA AATCCCCATG	780
	TTGATATAGT TGTTTAATGA AAAATGCGCC TAATTCCGTT TCATTATTAA AGAAATATTG	840
40	TTTGTTAGCA TAGTAAACAA TAATTTTTTG TACTTCTGGT TTGCCATCCT TGTAAGAAAA	900
	ATACTTTTCT AATTTTGTGT CACCTTCTGG ATTATAGAAA AATTCACATA ATGTTTGTTG	960
45	TTTATCAACA AGAATCCTAC TACAACCTAA AAAGCCACGC ACATCATAAA AATCACGTTT	1020
	TACTTtTCGT CTTTGACTAT CAAAATGATT TACATAATCT AATATACGAT ATTTAGGATC	1080
	TTGAAAATGG GCATACATTA AGAAACGCTC TTGATCATAT ATTCTAAAGT CATGACTATT	1140
50	TTCAACATGT TTAAAGTAT AATGACATTC ATCAGTCCAA TACGACAACC AGTCAAATGG	1200
	TTCAATTGCGT TCTAAATATG TTGCTTCTTG GAAGAAATCA TACATATTAA TATAGTCAGA	1260
55	ACTAGTAATA TAATTTTGGG CATTTCTATA TAAATATCTA TTCCATGACA GAAATACACA	1320

	CCCAGTTAAA TTAACACCTA AACTATTACC TACAAAATAA TTCATTTACA ACACCACTTA	1440
	TATCTATTTT TTATAATTAT ATCACATAAT ATTTAATTAC TTCTTTTAAC TGAAGATGT	1500
5	GTTTATTTAT AAAACAACAA ATTTTGATAT TTATAATGAT AGTAGTTATT CAATCACTAC	1560
	GACCCaATAT ATCATkGTAG AGCTTAGGAT ATTGATTTAT GACTCAGGCA CATCAAATGa	1620
	GAGGATTTAT AAARGAGATA TACAACCTA GAAGGTATAA TAAAAACGCG CAACTAATGT	1680
10	TACGCGTTTG AATTAATCAT ATGATATTAT TTGCGATACT TTAATTTAGC GAAAgCATCA	1740
	TGTTGATGGA TAGACTCTTC ATTACGACAT TCGATATCGA AACCGTCTAA CCAATCAAAT	1800
15	TCAACTAAGT CCGCGGCAAT TAAACGAATT AAGTCTTCGA CAAAACGTGG ATTTTCATAT	1860
	GCACGCTCTG TCACACGTTT TTCATCAGGA CGTTTTAAAA TAGGGTATAG AATTGAACTT	1920
	GCATTAGCTT CCATTGCATC TAAAATTTTA TTTTATAGT CATCAACTAT GTCTTGATCT	1980
20	TTATTAATAT ATGTTTTAAC AGTGACAACA CCACGTTGGT TGTGCGCTGA ATACTCACTT	2040
	ATTTCTTTTG AACAAAGGGCA TAGCGTTGTG ACAGTTGCTT CAATAGTAAG TTCTTTACGT	2100
	GTAnCTTTAT CACCGTCAAT TGCTAATCCA TAAGTGACAT CGGCATTACC AACTGCTTTA	2160
25	ATATTTGTGG TTGGAATA GCGATCAAAG AACCATTTCC CAGAAACATC AACGCCTGCC	2220
	GCATTTTGTT TCATATTCGT TTGTAAAGTG CGTAACACCT GATAAAGTGT ATTAAATTCA	2280
30	AGTTCAATAC CATTATCATA GTGCTTTTCA AACTTTTCGA TTATACGGCT CATATTAATA	2340
	CCTTTTTCGT CTTTGTAA ACTTGTGAA AAATAAATG TGCCAGCTGT TTGATACTGG	2400
	TCAACAAGTA CAGGGTACAC TAAGTTTTTA ATACCAACTT CTTCTATTTC AAATAAAAAA	2460
35	TCTTTATGTG TACTTTGTAA ATCTGTCATT TCGTTCTTAG TAGTAGGTTT CGTGCCTTCA	2520
	ATAGGATCTA CGGAACCAAA GTGTTTCCAA CGACCTTCTC GTGTCGATAA ATCAAATTCA	2580
	GTCA ¹ TTTTT TCCTCCGTTA AGATTTAAAG TGATATGTCC AATATGGTTC GACTGTTAAA	2640
40	AAGCTGTGTT GTTTACCATC GATTTTCAGGA CTTGCTAATT GTTTTAAAAA TGGACCTGTT	2700
	TGAGAAGCAT GTGCTTCAAA TGCCTTAATT TTAAGTTCTT TAAATCTGT AATATCATT	2760
	TGAATATCAG GTTCTCCAAG AGCTTCGGTT GCATCATTAC TGAACGCAAC TAAAGTTAAA	2820
45	CGAGGGCGTT CTTCTTTAGG CATGCGTTCA ACCGTTTCGA TTACAGCGTC TGCTGTTGCT	2880
	TCGTGATCAG GATGTACTGC ATATCCAGGA TAAATGAAA TAATCAATGA TGGATTTGTA	2940
50	TCATCGATTA AAGATTTAAT CATACCATCT ATATGTTTAT AGGGTTCAAA TTCGACAGTT	3000
	TTGTCACGTA AACCCATTTT TCTTAAATCA GTAATACCGA TAACTTTACA AGCTTCTTCT	3060
	AGTTCACGCT CACGAATACT TGGTAATGAT TCGCGTGTG CAAATGGGGG ATTACCTAAA	3120

	TAATTTGCTA ATGTGCCTGC AGATGAGAAG GTTTCATCAT CAGGATGTGG AAATATTACT	3240
	AATACATGTC TTTCGTCAGT CATGTTGATG CCTCCTCTAT AAATTAAATG GTCGCTCACT	3300
5	AAATTGAAGT GCTGCAGCGA GTTGACCTTC GTAATTAAAA CCTGCAATTA AAAATTCATC	3360
	ATGCTCATTG ACCTCAAAAT GCGTTAGACC TTGTACATAA ACCCAACCAC CATTTGATAG	3420
10	TTTAAGACCA ATGCGATAAG GTTCTTTATT ACCACCTTTT AGTTGTGCAT GCGTATATGT	3480
	TATTTGTATG TTCTTTAAAA AAGTACCAGC ATTAAAAACA CGTTGATCGA AATGGTTCGC	3540
	ATAGGCCCCA TTGTGCGTTT CAACATGCAG ATACACAGGT TTATGTTCAA AAGAAGCAAG	3600
15	TAAATCTATA ACTTCTTGTT CTTTAATTGG TTCCAACACG TTCACTCCTT ACACTATCAA	3660
	TGTGTTTATC TTCTATTTT ACTAAAACT ATTCGATAAT TGTATACGAT TGCTCAATTA	3720
	TTTATAAATT AATTTTCATG AAGGGTAATT ACTCAGGATT ACGTAATCAT ACAGCATTAG	3780
20	TTTTTTACTT TAAAAATCA AAAATTTGTT GGAATTTGAA AAGTGTTAAA CATTAAAAAT	3840
	GATGCTATAT TAATGGTGTA TGAATGAATT CATAAGTTTT TAAATGTAT TAAATTTGTG	3900
	GAGGCATGTA AACAAATGAAA GTATTAACT TAGGATCGAA AAAACAAGCA TCATTCTATG	3960
25	TTGCATGTGA GTTATATAAA GAGATGGCAT TTAATCAGCA CTGTAAACTA GGTTTAGCAA	4020
	CTGGTGGTAC AATGACAGAT TTGTATGAAC AACTTGTTAA GTTGTTAAAT AAAAATCAGT	4080
30	TAAACGTAGA CAATGTATCC ACGTTTAAAT TAGACGAATA TGTAGGTTTA ACCGCATCAC	4140
	ATCCGCAAAG TTATCACTAT TATATGGATG ACATGCTTTT CAAACAATAT CCTTATTTTA	4200
	ATAGAAAGAA CATTATATT CCAAATGGAG ATGCCGATGA TATGAATGCG GAAGCGTgCA	4260
35	AAATATAATG ACGTTTTAGA ACAACAAGGT CAACGTGATA TTCAAATTTT AGGTATTGGT	4320
	GAAAATGGTC ATATTGGATT TAATGAACCT GGTACGCCGT TTGATAGCGT TACTCATATC	4380
	GTTGATTTGA CTGAAaGTAC TATTAAGGCT AATAGTCGAT ATTTTAAAA CGAaGATGAT	4440
40	GTTCCAAAGC AAGCCATTTC GATGGGACTT GCTAATATTC TTCAAGCCAA ACGTATCATT	4500
	TTACTCGCAT TTGGTGAAAA GAAACGTGCT GCTATTACAC ATTTATTAAA TCAGGAAATT	4560
45	TCTGTTGATG TTCCAGCCAC ATTACTTCAC AAACACCCGA ATGTTGAGAT ATATTTAGAC	4620
	GACGAAGCTT GCCCGAAAAA TGTTGCGAAA ATTCATGTCG ATGAAATGGA TTGATTGCAA	4680
	TGTTTAATTA AGAAATGCCT CGGGAAAGGT TCCAATAGAA AGATAAAAAG CATTGGAAGG	4740
50	ATGATTTTTA GTGGAATTAC AATTAGCAAT TGATTTATTA AACAAAGAAG ACGCGGCTGA	4800
	GTTAGCAAAT AAAGTAAAAG ATTATGTAGA TATCGTAGAA ATCGGTACGC CAATCATTTA	4860
55	CAACGaAGGT TTACCAGCAG TTAAACATAT GGCAGACAAC ATTAGTAATG TAAAAGTATT	4920

CGCGGATGTA ATTACAATAC TAGGTGTTGC AGAAGATGCA TCAATTAAAG CAGCTATTGA 5040
 AGAAGCTCAT AAAAATAATA AACAATTACT AGTTGATATG ATTGCTGTTT AAGATTTAGA 5100
 5 AAAACGTGCA AAAGAACTAG ATGAAATGGG TGCTGATTAT ATTGCAGTAC AACTTGGTTA 5160
 TGATTTACAA GCAGAAGGGC AATCACCATT AGAAAGTTTA AGAACCGTTA AATCTGTTAT 5220
 TAAAAATTCT AAAGTTGCAG TAGCAGGTGG AATTAAACCA GATACAATTA AAGATATTGT 5280
 10 CGCTGAAAGT CCTGATCTTG TTATTGTTGG TGGCGGAATC GCAAATGCAG ATGATCCAGT 5340
 AGAAGCTGCG AAACAATGTC GCGCTGCAAT CGAAGGTAAG TAATATGGCT AAATTTAGTG 5400
 ACTATCAATT AATTCTAGAT GAATTAAAGA TGAATTTGTC ACATGTTGAA GCGGATGAGT 5460
 TTTCACTTTT TGCATCCAAA ATACTACATG CTGAACATAT ATTTGTAGCT GGCAAAGGAC 5520
 GTTCAGGATT CGTGGCGAAT AGTTTTGCAA TGCGCTTAAA TCAGCTCGGC AAACAGGCAC 5580
 20 ATGTTGTTGG AGAATCAACG ACACCTGCGA TTAAGTCGAA TGATGTATTT GTAATTATCT 5640
 CTGGTTCAGG TTCCACGGAA CATTTAAGAT TATTAGCAGA CAAAGCAAAA TCAGTAGGTG 5700
 CTGACATCGT ATTAATTACT ACAAATAAAG ATTCTGCAAT AGGCAATCTA GCTGGGACGA 5760
 25 ACATCGTTTT GCCTGCAGGT ACAAATATG ATGAACAAGG CTCGGCACAA CCATTAGGAA 5820
 GTTTGTTTGA ACAAGCATCT CAATTATTTT TAGATAGTGT TGTAATGGGA TTGATGACTG 5880
 AAATGAATGT TACGGAACAA ACGATGCAAC AAAATCATGC TAATTTAGAA TAAAATAAAG 5940
 30 ATAGTCGATA ATATGATGCC TAGGCAGAAA TATTATCGAT TATTTTTTTA TTTAAATAAT 6000
 AAATTATAGT ATAATATCAA TAATAACGA ATAGGGGTGT TAATATTGAA GTTTGACAAT 6060
 TATATTTTTG ATTTTGATGG TACGTGGCA GACACGAAAA AATGTGGTGA AGTAGCAACA 6120
 35 CAAAGTGCAT TTAAAGCATG TGGCTTAACG GAACCATCAT CTAAAGAAAT AACGCATTAT 6180
 ATGGGAATAC CTATTGAAGA ATCATTTTTA AAATTAGCAG ACCGACCATT AGATGAAGCA 6240
 40 GCATTAGCAA AGTTAATCGA TACATTTAGA CATACATATC AATCTATTGA AAAGGACTAT 6300
 ATTTATGAAT TTGCGGGTAT AACTGAAGCC ATTACAAGTT TGTATAACCA AGGGAAAAAA 6360
 CTTTTCTGG TGTCTAGTAA GAAGAGTGAT GTATTAGAAA GAAATTTATC GGCTATTGGA 6420
 45 TTAAATCACT TGATTACCGA AGCTGTTGGA TCCGATCAAG TAAGTGCATA TAAACCAAAT 6480
 CCTGAAGGCA TACACACAAT TGTGCAACGC TACAATTTAA ATAGCCAACA AACGGTGTAT 6540
 ATGGGTGATT CAACGTTTGA TGTGAGATG GCACAACGTG CTGGTATGCA ATCTGCAGCT 6600
 50 GTCATTGGG GTGCACATGA TGCAAGGTCA TTAATTCATT CAAATCCGGA TTTTATTATT 6660
 AATGATCCAT CAGAAATTAA TACCGTATTA TAAACTTGT TAAACAGAG AATACCATGG 6720

55

	ATTTAAAATA AATATTTATT AAACATTATG AATTTTTTAAA GAGTAATGTC TGACTCGTTG	6840
	ATAATTTATT TTGTAAAAA TAAATTAAAG TAATGACAAA GTTATTGAAG TAAATTGAGT	6900
5	ATAAACATTT AAATACGATG TCGAAAATGG CGATAGCATA TCACTTACAT GAAGTTGTGT	6960
	GctATCGCTA TTTTtagTTA TAATTCCAAA AAGTTAATCG TTCGATGATT TAAGAATTAT	7020
10	TATTGTTTAA TTCAAATGTA TGAGGGTATA AAATCATTGA ATTTAATTCG ATAAAGCGAA	7080
	ATTTTTGAAC AAACATACTT TTGTATTTAT ATAAAAGTTT AAATTCTTAT AAATTTGACA	7140
	AAACTAATTA ACTCCGTATA ATTATGAAAC ATACAAGAGG GAGTGTATGA ATTCATGGAT	7200
15	TTTAATAAAG AGAATATTAA CATGGTGGAT GCAAAGAAAG CTAAAAAAC CGTTGTTGCA	7260
	ACCGGTATCG GTAATGCAAT GGAATGGTTC GATTTTGGTG TCTATGCATA TACAAC TGCG	7320
	TACATTGGAG CGAACTTCTT CTCTCCAGTA GAGAATGCAG ACATTGACA AATGTTGACT	7380
20	TTCGCAGCAT TAGCCATTGC GTTTTTATTA AGACCAATTG GTGGTGTCTG ATTTGGTATT	7440
	ATTGGTGACA AATATGGACG TAAAGTTGTA TTAACATCTA CAATTATTTT AATGGCATTT	7500
	TCAACATTAA CCATTGGATT ATTGCCAAGC TATGATCAAA TTGGACTTTG GGCACCAATA	7560
25	CTATTATTGC TTGCAAGAGT ACTACAAGGG TTTTCAACAG GTGGAGAGTA TGCGGGGGCA	7620
	ATGACATATG TTGCCGAATC ATCTCCAGAT AAGCGTCGTA ACTCATTAGG TAGTGGACTA	7680
30	GAAATTGGGA CATTATCAGG TTACATAGCT GCTTCAATTA TGATTGCTGT ATTAACATTC	7740
	TTTTTAACAG ATGAACAAAT GGCATCATTT GGTGGAGAA TCCCATTCTT ACTCGGTTTA	7800
	TTCCTAGGAT TATTCGGCTT ATATTACGT CGTAAGCTGG AAGAATCACC AGTTTTCGAA	7860
35	AATGATGTTG CAACACAACC AGAAAGAGAT AACATTAACT TTTTACAAAT CATCAGATTT	7920
	TATTACAAAG ATATATTTGT ATGTTTTGTA GCTGTTGTAT TCTTCaATGT TACAACTAT	7980
	ATGGTAACTG CATATTTACC AACCTATTTA GAACAAGTTA TTAAATTAGA TGCAACGACA	8040
40	ACAAGTGAT TAATTACTTG TGTCATGGCA ATAATGATTC CATTAGCATT AATGTTTGGT	8100
	AAGTTAGCGG ATAAAATAGG TGAAAAGAAA GTATTTCTAA TTGGTACTGG TGGGCTAACA	8160
45	TTATTCAGTA TCATCGCATT TATGTTATTA CATTCACAAT CATTGTTGT AATAGTAATC	8220
	GGTATATTTA TATTAGGATT TTTCTTATCA ACTTACGAAG CGACAATGCC AGGGTCGTTA	8280
	CCAACGATGT TTTACAGTCA TATAAGATAT CGAACTTTAT CAGTAACATT TAATATCTCT	8340
50	GTTTCGATAT TTGGTGGTAc GaCGCCATTA GtKGCaMCaT GGTTaGTTAC GAAAAC TGGA	8400
	GATCCATTAG CmCCTGCGTA TTATTTAACA GCAATCAGTG TTATTGGCTT TTTAGTTATT	8460
55	ACATTCTTAC ATTTAAGTAC AGCAGGAAAA TCTCTAAAAG GTTCGTATCC AAATGTAGAT	8520

	GAACGTAAGA ATTAGAGATT TTAATaAAAA GTATAAATCA ATCGTATATA AGCACTTTAA	8640
	AGCTAGTAGG TTCTGCTAAC TTTAAAGTGC TTTTAAATT GAGAACTGTA ATTAGCCGTA	8700
5	ATAAAGTTTT TGTATATACA TAAACCCCCA CTGCAATGAT TATCGCAATG GGGGAAAGAG	8760
	GGGACTTAAA GCATATGTTT AGCTTTGAAT ACTTAAATTT CTCTTGCTAT TGAAATGTTA	8820
	GGATGTAAAT ATGTCTTAGA GTATTTTGTC CAACGCAATT AATATTGAGA CTCTAACCTT	8880
10	CAATATTATT ATAGAGAACA CAAACTTAAA TAGATTGGGT GACTTATTTG TGTCAATTAT	8940
	TGCGATTGCG ATAACCTCTT TTCTCTATAT ACATATAGTA ACGTCTTATC TAATAAAAAA	9000
	CATGGTACTA CAGTATCAAA TTTATCTAGG GCTTAAAGTTT GATTTTATA ATAGGCAGGT	9060
15	TTACCTGATA AAAATACTTA TTCATTATAT AATGTTAACA ATATGTATTT TAAAGTTTAC	9120
	ATTGAGTGAG GGATATTGAT GAACGTAATT TTAGAACAGT TGAAAACACA TACTCAAAT	9180
20	AAACCTAATG ACATAGCATT ACATATCGAT GATGAAACAA TTACATATAG TCAACTAAAT	9240
	GCCCCATCA CTAGCGCagT TGAATCTTTG CAGAAATATT CACTTAACCC TGTCGTTGCT	9300
	ATTAATATGA AATCACCGGT GCAAAGTATT ATTTGTTATT TAGCTTTGCA TCGTTTACAT	9360
25	AAAGTGCCTA TGATGATGGA AGGTAAATGG CAAAGTACTA TACATCGTCA ATTGATTGAA	9420
	AAATATGGTA TTAAAGATGT AATTGGAGAT ACAGGTCTCA TGCAGAATAT AGACTCACCG	9480
	ATGTTTATTG ATTCAACGCA ATTACAGCAC TACCCCAATT TATTACATAT TGGTTTTACT	9540
30	TCAGGGACAA CTGGACTGCC AAAAGCATAT TATCGTGATG AAGATTCATG GTTGGCTTCT	9600
	TTTGAAGTTA ATGAAATGTT GATGTTAAAA AATGAAAATG CAATAGCAGC CCCTGGACCA	9660
	CTATCGCACT CGTTAACATT ATATGCGTTA TTGTTTGCTT TAAGTTCCGG TCGTACTTTT	9720
35	ATAGGACAGA CCACTTTTCA TCCTGAAAAG TTACTIONAATC AATGTCATAA AATATCATCA	9780
	TACAAAGTTG CTATGTTTCT TGTTCCAACG ATGATTAAAT CATTATTGTT AGTTTACAAC	9840
	AATGAACATA CAATCCAATC ATTTTTTAGC AGTGGAGATA AGCTGCATTC TTCTATTTTT	9900
40	AAAAAGATAA AAAATCAAGC AAATGACATA AATTTGATTG AATTTTTTGG TACATCGGAA	9960
	ACCAGTTTTA TCAGCTATAA CTTGAATCAG CAAGCACCAG TTGAATCAGT AGGTGTGCTA	10020
45	TTTCCAAATG TGGAATTGAA AACAACGAAT CACGATCACA ATGGTATAGG AACTATTTGT	10080
	ATAAAAAGTA ATATGATGTT TAGTGGCTAT GTAAGTGAAC AATGTATAAA TAATGATGAA	10140
	TGGTTTGTTA CTAATGATAA TGGCTATGTA AAAGAGCAGT ATTTATATTT AACGGGACGT	10200
50	CAACAGGATA TGTTAATTAT TGGTGGTCAA AATATATATC CAGCACATGT TGAACGCCTT	10260
	TTAACGCAAT CTTGAGCAT TGATGAAGCA ATTATCATCG GTATTCCAAA TGAGCGTTTT	10320

55

	CAATTTTAA AAAAGAAAGT GAAaCgnTaT GAAATTCAT CGATGATTCA TCATGTAGAA	10440
	AAGATGTATT AACTGCAAG tGGTaAAAATT GCTAGAGAAA AAATGATGTC GATGTATTTG	10500
5	AGAGGTGAAT TATAATATGA ATCAAGCAGT CATAGTTGCA GCTAAACGAA CTGCATTTGG	10560
	GAAATATGGT GGCACTTTAA AACATTTAGA GCCaGAACAA TTGCTTAAAC CTTTATTCCA	10620
	ACATTTTAAA GAGAAGTATC CAGAGGTAAT ATCTAAAATA GATGATGTAG TTTTAGGTAA	10680
10	TGTTGTTGGG AATGGTGGCA ATATTGCAAG AAAAGCATTG CTTGAAGCGG GGCTTAAAGA	10740
	TTCAATACCT GGCCTCACAA TCGATCGGCA ATGTGGGTCT GGACTTGAAA GTGTTCAATA	10800
	TGCATGTCGC ATGATCCAAG CCGGAGCTGG CAAGGTATAT ATTGCAGGTG GTGTTGAAAG	10860
15	TACAAGTCGA GCACCTTGGA AAATCAAACG ACCGCATTCT GTGTACGAAA CAGCATTACC	10920
	TGAGTTTTAT GAGCGTGCAT CATTTGCACC TGAAATGAGC GACCCATCAA TGATTCAAGG	10980
20	TGCTGAAAAT GTGGCCAAGA TGTATGATGT TTCAAGAGAA TTACAAGATG AATTTGCTTA	11040
	TCGAAGTCAT CAATTGACAG CGGAAAATGT AAAGAATGGA AATATTTCTC AGGAAATATT	11100
	ACCTATAACC GTTAAAGGAG AAATATTCAA CACTGATGAA AGTCTAAAAT CACATATTCC	11160
25	GAAAGATAAC TTTGGCCGAT TTAAGCCCGT GATCAAAGGT GGGACCGTTA CCGCTGCGAA	11220
	TAGTTGTATG AAAAATGATG GTGCAGTTTT ATTGCTTATT ATGGAAAAAG ATATGGCATA	11280
	CGAATTAGGT TTCGAGCATG GTTTATTATT TAAAGATGGT GTTACGGTAG GTGTTGATTC	11340
30	TAATTTTCCT GGCATTGGTC CAGTACCAGC CATTTCCAAC TTAATAAAAA GAAATCAATT	11400
	AACGATAGAA AATATTGAAG TCATTGAAAT TAACGAAGCG TTCAGTGCAC AGGTAGTTGC	11460
	CTGCCAACAA GCTTTAAATA TTTCAAATAC GCAATTAAAT ATATGGGGTG GTGCATTAGC	11520
35	ATCAGGTCAT CCATACGGTG CAAGCGGTGC CCAATTAGTG ACTCGATTAT TTTATATGTT	11580
	TGACAAAGAG ACTATGATTG CATCTATGGG GATAGGGGGA GGTCTAGGAA ATGCAGCATT	11640
40	ATTTACTCGA TTCTAACCAG CGATTAAATG TGTCATTTTC TAAGGATAGT GTGGCTGCAT	11700
	ATTATCAGTG TTTTAACCAA CCTTATAGAA AAGAAGTACC ACCATTAAATG TGTGCGTCAT	11760
	TATGGCCAAA ATTTGATTTA TTTAAAAAAT ATGCAAATAG CGAACTGATT TTAACAAAAT	11820
45	CAGCAATTAA TCAAACCTCAA AAGATAGAAG TAGACACAAT ATATGTAGGG CATTTAGAAG	11880
	ATATTGAATG CCGACAGACT CGCAATATCA CACGTTATAC AATGGCTTTA ACATTAACTA	11940
	AAAATGATCA ACATGTCATA ACGGTtACAC AAACTTTTAT TAAGGCGATG AAGTAGAGAT	12000
50	GGAGTTTAAT GAGATATGGA TAAATGAATA TTTGGCGCTC GTAAATGATG ATAATCCAAT	12060
	ACATAATGAG ATTGTGCCAG GACAATTAGT GAGTCAAATG ATGCTGATGG CTATGTCATT	12120

55

	ATTCATTGAA CAACACGAAC ACGAAATTAT AGCAATTAAT GACGATGGAG AGATTAAAAT	12240
	AAAAATTTCT TTGAGCACAA AAAAATAACC GATATTAGCT GCATGAACGC ATATTAATTA	12300
5	GGAGATGAAA GGACAGCTAA TATCAGTTAT GTATTGTTAT TATTATTGGG AACAGAGATG	12360
	AATATAGGTT ACGTTTCTTT CTTTGCACGG GGATGCATTA ATCTAAAATA ATAATAACAA	12420
10	CTATATCAAT GTTTAATAAA TTCTGGATTA TTGGAACGAT TAGTCAATTT AACTAACTTT	12480
	CATATGATCT ATATCGTCTT GTAATAAAGA GAGCAATTTG AATATTTTCAG TATCACTAAA	12540
	TGAATCGTCA CATTTAATTG AAACATGCTG AAACGTTTTG GTTATAATTT CATAAACTGG	12600
15	TGCGCCTTCA TGGTGATACT GTCGATAAAT AATCATAACC TATATTACCT CCTTTGCTAC	12660
	TCTATGGTTA TATTATAAAT AACATTTTTA TGTGTGACAT CAACCTTAAG TATCAACTTT	12720
	TTATCAGACA TAGAACGTAT GATTTACTAA GACTATTTAT GTATAAAAGT TCTAAATAAA	12780
20	TATATATTTA TAGAGTCGCC TGGCAGTCAT TTGGGAAATA TAACATATAT GATTAGAGAG	12840
	GCATCTATCG CAAAAGAATG ATAATGATAG AGGTATTGAG CATATAGATG AGTTTAAGTT	12900
	CATCTTGAAA ATAAAGGGTT ATTTAGTCAT AGATGTAGAT GTATAGGAAA TATTTGTATG	12960
25	TATTGTTTGA TATGTATGAA ATTTTCAATA AAAGCTAATA ACGCTTATAT GTAACTTTCA	13020
	AATTTAAATT ATATACAGAG CATGATGATT ATAAAAAAT AACCACATCA CATAAATTGA	13080
30	GTTCATACCC AATTTAAGTG GTGTGGCTAA TAATGTTGAT TTATAGATGA ACCGCCTAAT	13140
	CGTTAAACCT CTGTTACTTC AACATCGATA TGTTCATAC GGTGTATGC ACCGTGATCC	13200
	ACAGGACCAA CAAAATCAAT CATTTTCCAA CCGTTTTTAA TAGCAGAAGC GACGAAAGCT	13260
35	TTGCGCTTAA TCACAGCTTC TTTCGGTGAC TTACCGTTAG CTAAATATGC AGTTGTTGCC	13320
	GCAGCAAATG TACAACCAGC ACCATGGTTA TAACTTTGTT GGAACATGTC TGTGTTAGT	13380
	TGATAAAATG TTTGACCATC ATAGTATAAG TCATACGATT TATCTTGATC TAAAGCTTTG	13440
40	CCACCTTTAA TGATGACATG CTGTGCGCCT TTATCAAAGA TAATTGTTGC AGCCTTTTTT	13500
	ATATCTTCAA TTGAATTTAA TTTACCTAAT CCTGATAATT GACCCGCTTC AAATAAGTTT	13560
	GGTGTCACTA CCGTTGCTTT AGGTAGTAAA TATTTAATCA TCGCCTCAGT ATTTCCAGGA	13620
45	TTAAGCACTT CATCTTCGCC TTTACAAACC ATGACAGGAT CTACTACAAA ATATTGTGCA	13680
	TTAGATGCCT CATATACTTC TCCAGCACGT TTGATTATCT CCTCAGTACC TAACATACCT	13740
	GTTTTAATAG CATCAGGTCC GATTGATAAA GCCGTTTCAA GTTGTTTTTC AAATACATCC	13800
50	ATTGGTAATG GTGTAACATC GTGTGACCAT GTATCTTTAT CCATAGTAAC GATGGCAGTT	13860
	AAAGCGACCA TGCCATACGT ATCTAATTCT TGGAACGTTT TCAAATCTGC TTGCATACCT	13920

55

	CACTCCTACA TAATAATATT GTATTCATCA TATCATTTTT AACCTAATTG AAAAATATTA	14040
	AGCATTCAAT ATTTGATGAT TGTGAAATG AATCATTTCAT ACTATTGTAA CTTTGGAAAA	14100
5	TGTCATTAC TTTAGATAAG TGTGATATGT TAAAATATGT CCTGAGGTGA GATTGAATGG	14160
	AATGGTCGCA AATTTTTCAT GACATAACAA CGAAACATGA CTTTAAAGCT ATGCATGATT	14220
	TTTTAGAAAA AGAATATTCG ACTGCAATCG TATACCCTGA TAGGGAAAAT ATATATCAAG	14280
10	CGTTTGATTT AACACCGTTT GAAAATATCA AAGTTGTTAT ATTAGGACAA GACCCGTATC	14340
	ATGGTCCAAA CCAAGCACAT GGATTAGCAT TTTCAGTGCA ACCTAACGCA AAATTCCCTC	14400
	CATCTTTACG TAATATGTAT AAAGAATTAG CAGATGATAT TGGATGCGTT AGACAAACAC	14460
15	CGCATTTACA AGATTGGGCA AGAGAAGGCG TCTTGTTATT GAATACAGTT TTAACCGTAA	14520
	GACAGGGTGA AGCAAATTCT CATCGTGATA TTGGTTGGGA AACATTTACT GATGAAATTA	14580
20	TTAAAGCAGT GTCTGATTAT AAAGAACATG TTGTCTTTAT TTTGTGGGGG AAACCTGCAC	14640
	AGCAAAAAAT AAAGCTTATC GATACATCTA AACATTGTAT TATAAAATCA GTGCATCCTA	14700
	GTCCACTGTC TGCATATAGA GGATTCTTTG GATCAAAACC GTATTCCAAA GCGAATGCCT	14760
25	ATTTAGAGTC AGTAGGAAAA TCACCAATTA ATTGGTGTGA AAGTGAGGCG TAGATGTTGA	14820
	ATAGAGAAAC TTTAATAGCA CGAATTGAGC AAGAATTAGT ACAAGCAGAG CAGGCACAGC	14880
	ATGACCATGA CTTTGGAAAA CATATGTATG CCATACATAT ATTAACATCT TTATATGCTT	14940
30	CAACATCAAA TACACCACAT ATTGGTGAAC AACAAATGAA TCGTCGTATT GCTAACCATA	15000
	ATCAAATGCC ACAATCACAA ATAACGCAGC CAACTCATCA AGTGACAGTT GCTGAAATTG	15060
	AAGCGATGGG TGGTAAAGTA AATACGCATT CAGCACATCA TCATAATAAG TCATATTAC	15120
35	AACCTTCAAA CCAACAACAA AGATTAGCGA CAGATGATGA CATTGGCAAT GGTGAATCCA	15180
	TATTGATTTT TTAAAAAGCA ACAATGAAAC ATAATTACTT AATAGCTTGT TAAGTATGTA	15240
40	GGTTAATAAT CAAGACGCAT ATACTTTTAT TCGAGTGTTT GGATTTAAAC ATTTATTAAT	15300
	ACTGAATTAT ATAAGGAGAG GTAGCAATGA AATTATTTAT TATTTTAGGT GCATTAAACG	15360
	CGATGATGGC TGTCGGTACA GGTGCATTTG GTGCGCATGG TTTACAAGGA AAAATAAGTG	15420
45	ATCACTATTT ATCAGTATGG GAAAAAGCAA CGACGTATCA AATGTACCAT GGCTTAGCAT	15480
	TATTAATTAT AGGTGTAATT AGTGGTACAA CTTCAATCAA TGTTAACTGG GCTGGCTGGT	15540
	TAATATTTGC TGGTATTATT TTCTTTAGTG GATCATTATA TATTTTAGTA TTAACCAAA	15600
50	TTAAAGTTTT AGGTGCGATT ACGCCAATTG GTGGCGTATT GTTCATCATT GGATGGATAA	15660
	TGTTAATCAT TGCGACATTC AAATTGCTG GTTAAATTTT AAAACTTTAG ATTACCTATG	15720

55

	TGGGTATAGA ATACCTTCGA GGTGAGTTTT TATTTATGGA AAAAAAGAAT AAGCAAATAG	15840
	ATAGAGGCGA TTAAAAACAA AACCTATCTG AAAAGTTTGT ATGGGCGATT GCATATGGTT	15900
5	CATGTATCGG ATGGGGCGCA TTCATCTTAC CAGGAGACTG GATTAAGCAG TCAGGTCCGA	15960
	TTGCAGCATC AATTGGTATA GTTATTGGTG CATTATTAAT GATATTAATT GCGGTTAGTT	16020
	ATGGGCGATT AGTAGAGAGA TTTCCAGTAT CAGGGGGCGC GTTTGCCTTT AGTTTCTTAA	16080
10	GTTCGGCAG ATATGTGAGT TTCTTCTCAT CATGGTTTTT AACTTTTGGT TATGTCTGTG	16140
	TCGTTGCTTT AAAtGCGACC GCATTCAAGT TACTAGTTAA ATTCTTATTG CCAGATGTCT	16200
	TAAATAATGG GAAACTATAC ACCATTGCGG GCTGGGACGT TTATATTACG GAAATCATT	16260
15	TTGCGACCGT ATTACTACTT GTATTCATGC TAGTAACGAT TCGTGGCGCA AGTGTATCTG	16320
	GATCATTACA ATATTATTTT TGTGTGGCGA TGGTAATCGT CGTATTATTG ATGTTCTTTG	16380
20	GTTCACTCTT TGGTAATAAT TTTGCACTTG AAAATTTACA ACCGTTAGCT GAACCTAGCA	16440
	AAGGATGGTT AGTGTCTATT GTGGTTATTG TATCCGTGGC ACCATGGGCA TATGTTGGAT	16500
	TTGATAATAT TCCACAAACA GCAGAAGAGT TTAACCTTGC ACCAAACAAG ACATTTAAGC	16560
25	TTATCGTGTA CAGTTTATTA GCAGCATCAT TAACTTATGT TGTCAATGATT TTATACACTG	16620
	GTGTTTATC AACAAGTCAT CAAAGTTTAA ATGGGCAGTT GTGGTTAACA GGTGCTGtTA	16680
	CACAAACAGC ATTTGGTTAT ATTGGATTAG GTGTATTAGC AATTGCAATT ATGATGGGTA	16740
30	TATTTACTGG TTAAATGGA TTCTTGATGA GTTCAAGTCG CTTGTTATTT TCTATGGGAC	16800
	GTTCAAGTAT TATGCCAACA ATGTTTAGTA AATTACATAG TAAATACAAA ACACCATATG	16860
	TCGCAATCAT ATTCCTAGTA GGAGTGTCTG TAATTGCACC TTGGCTAGGA AGAACTGCAT	16920
35	TGACTTGGAT TGATAGATATG TCATCTACTG GTGTATCCAT TGCCTACTTT ATTACATGTT	16980
	TGTGTGTCAGC GAAATTATTC AGTTATAACA AACAAAGTAA TACGTATGCA CCGGTTTACA	17040
40	AAACGTTTGC TATTATCGGC TCATTTGTAT CATTCATTTT CTTAGCGTTG TTATTAGTGC	17100
	CAGGTTCTCC TGCAGCACTG ACTGCACCGT CTTATATTGC ATTACTTGA TGGTTAATCA	17160
	TCGGTTTAAT ATTCTTTGTG ATTCGATATC CTAAATTGAA AAATATGGAT AATGATGAAT	17220
45	TAAGTCGCTT GATTTTAAAT AGAAGTGAAA ATGAAGTTGA TGATATGATT GAAGAACCTG	17280
	AAAAAGAAAA AACTAAATAA TAAAAGAATC GCACAATAAA CCTTCTTCAT TCGGAGGCGT	17340
	ATCGTGCGAT TTTTGTATT ATAAATTGAC ATTTAAGACG AGGCAGCTGA ACCTTATATA	17400
50	TAATTGCTAA GAGTTAGGGC TGAGCCATTT CTAACAAATA TTTATAATCG TTTAAAAGAT	17460
	TTACGAACC CAGAAACAAT TAATTTGGAA ATTTGGTCGG CGAATAATAA ACCTAATGCG	17520

55

AAGACTAAAT TTTTGTAGC ATCGTATGCT AAGCCACCAG GACTAATGG AATGATACCC 17640
 GTTACCATAA AAATGATGGC AGGTTCTTTT TGTTTACGAG CCATATAATG ACTTAACAAG 17700
 5 CCTAATGCTA AACTACCAA GAACTAGAG TATATAGTGT GCACATTAAA GCCGTTGAAG 17760
 AATAAGGTGT AAACCATCCA TCCACACGTA CCAACGAAAC CACATGATAG ATATAATTTT 17820
 CTAGGTGCAT CAAAATGAC GCAGAA 17846

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5544 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

ATTGACACTT GGTGAAAGTA ATATCGCCGC GCTATTTTGG CAAAATGGAC ACTTAGAACC 60
 TGAGTTACAA GATGAACAGC CAATTAATAT ATTAGGATCT GkTCAAATCA ACGAATGGAA 120
 25 TGGTAATCAA TCACCGCAA TAATTATTCA AGATATTGCG ATGAATGAAC AGCAAATATT 180
 AGATTATAGA AGTAAGCGAA AAAGTTTACC TTTTACAGAA AATGATGAAA ATATTGTCGT 240
 GCTTATTCAT CCTAAAAGTG ATAAAGTAAA TGCGAATGAA TATTATTATG GTGAAGAAAT 300
 30 TAAACAACAA ACTGATAAAG TAGTATTAAG AGATTTACCA ACGTCAATGG AAGACTTGTC 360
 TAATTCCTTG CAACAACCTGC AATTTTCTCA ACTTTATATA GTTTTGCAAC ATAATCATTC 420
 GATTTACTTC GATGGTATAC CTAATATGGA TATTTTAAA AAGTGTATA AAGCATTAAT 480
 35 AACTAAACAA GAAACAAATA TCCAGAAAGA GGGTATGTTA TTGTGTCAAC ATTTAAGTGT 540
 GAAACAGAT AACTTTAAAT TCATGTTGAA AGTTTCTTA GACTTAAAT TTGTAACACA 600
 40 AGAAGATGGT TTAATTCGAA TCAATCAACA ACCTGATAAA AGATCGATTG ATTCCAGCAA 660
 AGTATATCAA TTAAGACAAC AACGTATGGA TGTTGAAAAG CAATTATTAT ATCAAGATTT 720
 TTCAGAAATA AAAAATTGGA TAAAGTCACA ATTGTCGTGA GCAATTTAGG AGGAAATATT 780
 45 AATGGATTTA AAGCAATACG TATCAGAAGT TCAAGATTGG CCGAAACCAG GTGTTAGTTT 840
 CAAGGATATT ACTACAATTA TGGATAATGG TGAAGCATAT GGCTATGCAA CAGATAAAAT 900
 TGTAGAATAC GCAAAAGACA GAGATGTTGA TATCGTTGTA GGACCTGAAG CGCGTGGCTT 960
 50 TATCATTGGC TGTCTGTAG CTTATTCAAT GGGGATTGGC TTTGCACCTG TTAGAAAAGA 1020
 AGGGAAATTA CCTCGTmAG TCATTCGTTA TGAGTATGAC CTAGAATATG GTACAAATGT 1080

	ATTAGCTACT	GGTGGTACGA	TTGAAGCAGC	AATAAAATTA	GTTGAAAAAT	TAGGCGGTAT	1200
	CGTAGTAGGT	ATTGCATTTA	TAATTGAATT	GAAATATTTA	AATGGTATTG	AAAAAATTAA	1260
5	AGATTACGAT	GTTATGAGTT	TAATCTCATA	CGACGAATAA	TAAATAATAT	AATTTTATCA	1320
	AATGAAATCC	TTCATCAAAT	GTATAAGAAC	CAATGACTTA	ATTAAAAAAG	TTGTTTAAAGT	1380
	TTTCTTAACA	TGAGATGTTA	GGATTTTTTTA	TTTACTGAAA	ATGTTAGATG	ATTGAGCAIT	1440
10	ATACCTTAAT	AACATCGTTT	ATTTATTTCA	TAAATTGTAG	TATCATAGAA	CTAATATTTA	1500
	AAAAATGAAA	CAGTAGATTT	AGGTCGAATT	TTTGTAAGG	TTTTAAAAGT	AGGAATAGTA	1560
	TACAAATTAA	ACTCGCTCAA	GTAAAATTAA	TATTACGATT	AATGACGACA	GGATAAATAT	1620
15	TTATCGTCGA	CGGACGTATG	ATTGGTGTGG	GACAAATACT	ATTCAACAAG	AGTACCTAAA	1680
	TCATTGTTTA	AGGCGAAGTA	ATAAATATGA	ATGGGGTGTA	TCATATAATG	AACAACGAAT	1740
20	ATCCATATAG	TGCAGACGAA	tTCTTCACAA	AGCAAAATCA	TATTTGTCAG	CAGATGAATA	1800
	TGAGTATGTT	TTAAAAAGCT	ATCATATTGC	TTATGAAGCA	CATAAAGGTC	AGTTCCGAAA	1860
	AAACGGATTA	CCATACATTA	TGCATCCTAT	ACAAGTTGCA	GGTATTTTAA	CAGAAATGCG	1920
25	ATTAGACGGA	CCGACGATTG	TCGCAGGTTT	TTTGCATGAT	GTAATTGAAG	ATACACCGTA	1980
	TACATTTGAA	GATGTAAAAG	AAATGTTCAA	TGAAGAAGTT	GCTCGAATTG	TTGATGGTGT	2040
	GACGAAGCTT	AAAAAAGTAA	AATACCGCTC	AAAAGAAGAA	CAACAAGCTG	AAAATCATCG	2100
30	CAAGTTATTT	ATTGCGATTG	CCAAAGATGT	ACGCGTAATT	TTGGTGAAAT	TAGCAGACAG	2160
	ATTACATAAT	ATGCGTACCT	TGAAAGCCAT	GCCGCGCGAA	AAACAAATTA	GAATTTCTCG	2220
	AGAAACATTA	GAAATTTATG	CACCATTAGC	ACATCGTCTT	GGTATTAATA	CAATCAAATG	2280
35	GGAACTAGAA	GATACGGCTC	TTCGTTATAT	TGATAATGTG	CAATATTTTA	GAATAGTCAA	2340
	TTTAATGAAG	AAGAAACGTA	GTGAACGTGA	AGCGTATATC	GAAACGGCTA	TTGATAGAAT	2400
40	ACGTACTGAA	ATGGACCGAA	TGAATATCGA	AGGCGATATA	AATGGTAGAC	CTAAACATAT	2460
	TTACAGTATT	TATCGGAAAA	TGATGAAGCA	GAAAAACAA	TTTGATCAAA	TTTTTGATTT	2520
	GTTGGCGATA	CGTGTATTG	TCAATTCTAT	TAATGATTGT	TATGCGATAC	TTGGGTTGGT	2580
45	GCATACGTTA	TGGAAACCGA	TGCCAGGACG	TTTTAAAGAT	TATATTGCAA	TGCCTAAACA	2640
	AAATTTGTAT	CAGTCATTGC	ATACTACAGT	AGTAGGCCCA	AATGGAGACC	CGCTCGAAAT	2700
	CCAAATACGA	ACGTTTGATA	TGCACGAAAT	TGCTGAGCAT	GGTGTTCAG	CACACTGGGC	2760
50	TTACAAAGAA	GGTAAAAAAG	TAAGTGAAAA	AGATCAAAC	TATCAAAATA	AGTTAAATTG	2820
	GTTAAAAGAA	TTAGCTGAAG	CGGATCATAC	ATCGTCTGAC	GCTCAAGAAT	TTATGGAAAC	2880

55

	TGAGTTGCCA TATGGTGCTG TGCCGATTGA TTTTGCTTAT GCGATTCACA GTGAAGTAGG	3000
	TAATAAGATG ATTGGTGCCA AGGTGAATGG CAAAATTGTA CCAATTGACT ATATTTTACA	3060
5	AACAGGCGAT ATTGTTGAAA TACGTACTAG TAAACATTCA TATGGACCAA GTCGTGATTG	3120
	GTTGAAAATT GTTAAATCGT CTAGTGCCAA AGGTAAAATT AAAAGTTTCT TCAAAAACA	3180
	AGATCGTTCA TCTAATATTG AAAAAGGCCG AATGATGGTT GAAGCTGAAA TAAAAGAGCA	3240
10	AGGATTTAGA GTCGAAGATA TTTTGACAGA GAAAAATATT CAGGTTGTTA ATGAAAAATA	3300
	TAACTTTGCA AATGAAGATG ATTTATTGCG AGCTGTAGGA TTTGGCGGCG TGACATCCTT	3360
15	ACAGATTGTT AATAAATTAA CTGAAAGACA ACGTATTTTA GATAACAAC GTGCTTTAAA	3420
	TGAAGCACAA GAAGTTACGA AATCATTGCC TATTAAAGAC AACATCATTG CTGATAGTGG	3480
	TGTCTATGTA GAAGGTTTAG AAAATGTACT TATCAAGTTG TCAAAATGTT GTAATCCTAT	3540
20	ACCaGGTGAT GATATTGTAG GTTATATCAC CAAAGGTCAC GGTATTAAAG TACATCGCAC	3600
	TGATTGCCCA AATATTAAGA ACGAACTGA ACGACTAATT AATGTTGAAT GGGTAAATC	3660
	AAAAGACGCA ACTCAAAAT ATCAGGTTGA TTTAGAGGTA AtGCGTATGA CCGAAATGGC	3720
25	TTGTTGAATG AAGTACTACA AGCTGTTAGC TCGACAGCCG GCAATTTAAT TAAAGTTTCA	3780
	GGACGTTGAG ATATTGATAA AAATGCAATA ATAAATATTA GTGTCATGGT GAAAAACGTG	3840
	AATGATGTTT ATCGTGTGGT AGAAAAGATC AAACAACTTG GTGATGTTTA TACAGTAACA	3900
30	AGAGTTTGGA ACTAGAGGTG CAAAATATGA AAGTAGTTGT ACAAAGAGTT AAAGAAGCAT	3960
	CGGTGACGAA TGATACATTA AATAATCAAA TCAAAAAGG ATATTGTTTA TTAGTCGGTA	4020
35	TCGGTCAGAA CTCTACAGAG CAAGATGCAG ATGTAATTGC AAAGAAAATT GCTAATGCAA	4080
	GATTATTTGA AGATGACAAT AATAAATTAA ACTTTAATAT CCAACAAATG AATGGTGAAA	4140
	TACTATCAGT TTCACAATTT ACTCTCTATG CAGATGTAAA AAAAGGTAAC CGTCCAGGTT	4200
40	TCTCAAATTC TAAAAATCCT GATCaAGCGG TAAAAATTTA TGAGTATTTT AATGcaTGCG	4260
	CTACGAGCGT ATGGTCTTAC TGTGAAAACA GGTGAATTTG GAACACACAT GAATGTTAGC	4320
	ATAAATAATG ATGGTCCAGT CACTATTATT TATGAAAGTC AGGACGGCAA AATTCAATGA	4380
45	AAAAATAGA GGCATGGTTA TCTAAAAGG GTCTTAAAAA TAAACGTACT CTAATAGTAG	4440
	TGATTGCCTT TGTCTTATTT ATCATCTTTT TATTTTTATT GCTGAATAGC AATAGTGAAG	4500
	ATAGTGGGAA CATCACGATA ACTGAAAATG CTGAATTACG TACAGGTCCA AACGCTGCGT	4560
50	ATCCAGTCAT ATATAAGTT GAAAAAGGTG ACCATTTTAA AAAGATTGGT AAAGTAGGTA	4620
	AATGGATTGA AGTTGAAGAT ACATCCAGTA ATGAAAAGG TTGGATAGCT GGATGGCACA	4680

55

TAGTGCTTGA TCCTGGTCAT GGAGGTAGTG ACCAGGGTGC TTCAAGCAAT ACTAAATATA 4800
 AAAGTTTAGA AAAAGATTAT ACGTTGAAAA CAGCAAAAGA ATTGCAGCGT ACTTTAGAAA 4860
 5 AAGAAGGCGC AACTGTTAAG ATGACAAGAA CAGACGATAC ATATGTTTCA CTAGAAAATC 4920
 GTGATATCAA AGGCGATGCC TATTTGAGTA TACATAATGA TGC GTTAGAA TCATCTAATG 4980
 CAAATGGAAT GACaGTTTAT TGGTATCATG ATAATCAAAG AGCTTTAGCA GATACGTTAG 5040
 10 ACGCTACGAT TCAGAAGAAA GGTCTACTTT CTAATCGCGG TTCAAGACAA GAAAATTATC 5100
 AAGTGTTAAG ACAACAAAA GTTCCTGCTG TTTTATTAGA ATTAGGTTAT ATTAGTAACC 5160
 CAACTGATGA AACGATGATT AAAGATCAAT TACATAGACA AATTTTAGAA CAAGCAATTG 5220
 15 TTGATGGCCT TAAAATTTAT TTTTCTGCGT AGGGCTTGCA AAAATATGTG AAAGTAGTTA 5280
 TCATTGATAT TGAATTTTAT AACTAAAACC GTTAGTATTC TTGAAATGGT AAATGAAATA 5340
 20 GGTAGCAATC TAACTAAGAT TGTGTAGGAA TATAATCCAT AGACTGAAAG ATTATGCTGA 5400
 GTAGTTTATA TACATTGAAC ACAAGAAGAG GTGCTTTATG AAAAGTAAAG CCGTTAAACG 5460
 TACGTTaAAC GTTTTGAGTG GGTTTATTAA ATGCACGCTT ATAAAAAGTA ATGATGATTA 5520
 25 CAATTAGGCA TGT TTTTAA ACCA 5544

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
 30 (A) LENGTH: 1067 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AAAAGATTGC AAATATAAAT GGCATGTTTA ATATGTTAGA ACAACAAATC ATTCATAGCC 60
 40 AAGATATGGC TCATTTTAGA AGTGAATTTT TTTACGTCAA TCATGaGCAT CGAGAAAAC 120
 ATGAAGCACT CCTAATTTAT TACAAAAATA GTATCGACAA TCCTATTGTA GATGGTGCAT 180
 GTTATATTTT AGCCCTACCT GAAATTTTCA ATAGTGTTGA TGTTTTCGAA TCAGAGTTAC 240
 45 CATTTTCATG GGTATATGAT GAAAATGGCA TTACCGAAAC AATGAAATCA CTTAGCATT 300
 CATTACAATA TTTAGTTGCA GCAGCTTTAG AAGTAACTGA TGTGAATATA TTTAAGCCTT 360
 CAGGATTTAC AATGGGAATG AATAATTGGA ATATTGCTCA AATGCGAATC TTTTGGCAAT 420
 50 ATACAGCAAT TATTAGAAAA GAAGCACTAT AACATTAATA ATTAATTAGC TATAAAGATG 480
 ATTCACAACA ATCATCTTTA TAGCTTTTTT ATGTCTAATT ATTTTGTAGG AAAATmACAA 540

55

AATTTTATGT TTTCAAAAGT AAACAATCAA AAGATGTTAG AAGATTGCTT CTATATAAGA 660
 AAGAAAGTGT TTGTAGAAGA ACAAGGCGTC CCTGAGGAAA GTGAAATTGA TGAATATGAA 720
 5 TCTGAATCTA TTCACCTCAT TGGATATGAT AATGGACAGC CAGTTGCCAC TGCTCGAATA 780
 CGCCCTATTA ATGAAACAAC TGTCAAATA GAACGAGTAG CTGTGATGAA ATCACATCGT 840
 GGACAAGGAA TGGGTAGAAT GCTTATGCAA GCTGTAGAAT CATTAGCTAA AGATGAAGGT 900
 10 TTTTACGTAG CTAATATGAA TGCCCAATGT CATGCTATCC CATTTTATGA AAGTTTAAAC 960
 TTTAAATGA GAGGTAATAT ATTTCTTGAG GAAGGCATCG AGCATATTGA AATGACAAAA 1020
 AAGTTAACCT CGCTTAATTA AAAAAAGTTG TATCTATTTT AGAAACA 1067
 15

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18613 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 20

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

AAGACGtAtG ATAACAACAA TACgTGtAGT GAAAGATTTT AATCTACATA TTACTGACAA 60
 AGAATTCATT GTATTTGTTG GACCATCGGG ATGTGGTAAA TCAACAACAT TACGAATGGT 120
 30 TGCTGGACTA GAGTCTATCA CATCTGGAGA TTTTATATT GATGGGGAAC GCATGAACGA 180
 TGTGAACCA AAGAATAGAG ATATTGCGAT GGTATTTCAA AACTATGCAT TATATCCACA 240
 TATGACTGTT TTTGAAAATA TGGCATTGG GCTAAAGCTA CGTAAAGTAA ATAAAAAAGA 300
 35 GATTGAACAA AAAGTTAATG AAGCAGCTGA AATATTAGGA TTAAGTGAAT ATCTTGGTCG 360
 TAAACCAAAA GCGTTATCTG GCGGACAGCG TCAACGTGTT GCTTTGGGCA GAGCTATTGT 420
 TAGGGATGCG AAAGTCTTTT TAATGGATGA ACCATTATCG AATCTTGATG CGAAyTtCGA 480
 40 GTACAAATGC GCACAGAAAT ATTGAAATTA CATAAGCGAC TTAATACTAC GACAATTTAT 540
 GTTACACATG ATCAAAGTGA AGCATTGACG ATGGCTAGTC GAATTGTTGT TTTGAAAGAT 600
 45 GGCGACATTA TGCAAGTCGG CACACCTAGA GAAATATATG ATGCCCCTAA TTGCATATTT 660
 GTGGCGCAAT TTATCGGCTC ACCAGCAATG AATATGTTGA ATGCTACAGT TGAAATGGAC 720
 GGATTGAAGG TAGGAACACA CCATTTTAAA TTACATAATA AAAAAATTGA AAAGTTAAAA 780
 50 GCTGCTGGCT ACTTAGACAA GGAAATTATT TTAGGTATTC GAGCTGAAGA CATTATGAA 840
 GAACCAATAT TTATTCAAAC TTCTCCAGAG ACACAATTTG AATCTGAAGT AGTTGTATCC 900

55

	AAATTAGATT CAAGAACTCA AGTGATGGCG AACGACAAGA TTACACTAGC ATTTGATATG	1020
	AATAAGTGTC ACTTTTTTGA TGAAAAAACA GGAAATCGTA TCGTCTAAGG GGGAGTATTC	1080
5	ATGTCTAAAA TTTTAAATG TATCACGTTA GCCGTGGTAA TGTTATTAAT CGTAACTGCA	1140
	TGTGGCCCTA ATCGTTCGAA AGAAGATATT GATAAAGCAT TGAATAAAGA TAATTCTAAA	1200
	GACAAGCCTA ACCAACTTAC GATGTGGGTG GATGGCGACA AGCAAATGGC GTTTTATAAA	1260
10	AAAATTACGG ATCAATATAC TAAAAAACT GGCATCAAAG TAAAGCTTGT AAATATTGGT	1320
	CAAAATGATC AACTAGAAAA TATTTGCTA GACGCTCCTG CAGGAAAAGG TCCAGATATC	1380
	TTTTTCTTAG CACATGATAA TACTGGAAGT GCCTATCTAC AAGGCTTAGC TGCTGAAATC	1440
15	AAATTATCAA AAGATGAGTT GAAAGGTTTC AATAAGCAAG CACTTAAAGC GATGAATTAT	1500
	GACAATAAGC AACTAGCATT GCCAGCTATC GTTGAAACAA CCGCACTTTT TTATAATAAA	1560
20	AAATTAGTGA AAAATGCACC GCAACGTTA GAAGAAGTTG AAGCTAATGC TGCCAAACTA	1620
	ACTGATAGTA AAAAGAAACA ATACGGTATG TTATTTGATG CTAAAAATTT CTATTTTAAT	1680
	TATCCGTTTT TATTCGGCAA TGATGATTAT ATTTTCAAGA AAAATGGCAG TGAATATGAT	1740
25	ATTCATCAGC TAGGACTAAA TTCAAAACAT GTCGTCAAGA ATGCTGAACG ATTACAAAAA	1800
	TGGTACGACA AAGGGTATCT TCCTAAGGCA GCAACACATG ATGTCATGAT TGGTCTTTTT	1860
	AAAGAAGGAA AAGTAGGACA ATTTGTCACT GGACCGTGA ACATTAATGA ATATCAAGAA	1920
30	ACGTTTGGTA AAGATTTAGG AGTAACAACA TTACCTACAG ATGGTGGCAA ACCTATGAAA	1980
	CCATTTCTAG GTGTACGTGG TTGGTATTTA TCTGAATATA GTAAACATAA GTATTGGGCT	2040
	AAAGATTTAA TGCTGTATAT CACTAGTAAA GATACATTAC AAAATATATC AGATGAAATG	2100
35	AGCGAAATTA CTGGACGTGT TGACGTGAAA TCATCTAATC CAAATTTAAA AGTGTTTGAA	2160
	AAGCAAGCAC GTCATGCTGA ACCGATGCCT AATATTCCTG AAATGCGACA AGTTTGGGAA	2220
40	CCGATGGGCA ATGCAAGCAT ATTTATTTCA AATGGTAAGA ATCCTAAACA AGCGTTAGAT	2280
	GAGGCGACGA ATGATATAAC GCAAAATATT AAGATTCTTC ATCCATCACA AAATGATAAG	2340
	AAAGGAGATT AGTTATGACG AAACGTAACC CTAAATTAGC GGCATTATTA TCTGTTATAC	2400
45	CTGGTTTGGG ACAGTTTTAT AATAAAAGAC CCATTAAAGG GACGATATTT TTTATCTTTT	2460
	TCATCAGTTT TATTTCTGTT TTTTATAGCT TTTTAAATAT TGGTTTTTGG GGATTGTTCA	2520
	CATTAGGGAC AGTACCTAAG TTAGACGATT CTCGTGTCTT ACTGACACAA GGTATTATTT	2580
50	CTATCTTACT CGTTGCTTTC GCAATCATGC TATATATCAT TAATATTTTA GATGCATATC	2640
	GTAATGCTGA ACGATTTAAT CGCAATGAGG AAATAAAGGA TCCGAAGCGC GTATGGTGGC	2700

55

	TGTAGTTGTA TTTCCATTAA TAYYTATGTT TGGAGTAGCA TTTACAAATT ACAATTTATA	2820
	CAACGCGCCT CCGAGACACA CATTAGAATG GGTGGTTTAA GATAACTTTA AAACGTTATT	2880
5	CACAATTGGC GTTTGGCGTA AAACATTTTT CAGTGTATT ACTTGGACAT TAGTATGGAC	2940
	GCTTGTGCA ACGACACTTC AAATTGCATT AGGGCTGTTT TTGGCAATTA TTGTAAATCA	3000
	CCCTGTCGTC AAAGGTAAGA AATTTATCCG TACTGTGTTA ATCCTACCTT GGGCTGTACC	3060
10	ATCATTGTG ACAATTTTAA TATTTGTAGC GTTATTTAAT GATGAATTTG GTGCGATAAA	3120
	TAATGATATT TTGCAACCTT TATTAGGTGT AGCACCAGCA TGGTTAAGTG ATCCGTTTTG	3180
15	GGCAAAAGTG GCATTAATCG GCATTCAAGT ATGGCTTGGA TTCCCATTG TCTTTGCACT	3240
	GTTCACTGGA GTACTGCAAA GTATTTCATC AGATTGGTAC GAAGCAGCAG ATATGGATGG	3300
	TGCGTCTAGT TGGCAAAAGT TTAGAAACAT CACATTCCCG CATGTCATTT ACGCCACAGC	3360
20	GCCATTGTTA ATTATGCAAT ATGCAGGTAA TTTCAATAAT TTTAATCTTA TTTATCTATT	3420
	TAATAAAGGC GGTCCACCAG TGTCAGGGCA GAATGCTGGT AGTACAGATA TCTTGATATC	3480
	TTGGGTGTAT AATCTGACAT TTGAGTTTAA CAACTTCAAC ATGGGTGCAG TTGTGTCATT	3540
25	AATTATTGGA TTTATTGTTG CTATTGTCGC ATTTATTCAA TTCAGACGTA CAAGTACGTT	3600
	TAAAGATGAG GGAGGTTTAT AAGATGACAA AGAAGAAAAA CATATTAAAA GCAATCGGTA	3660
	TTTACAGTTT TATAGCGATG ATGTTTGTC TCAATTTTATA TCCACTACTG TGGACATTTG	3720
30	GCATTTCCCT TAATCCAGGT ACGAACTTGT ATGGTGCCAA AATGATACCA GACAATGCAA	3780
	CATTTAAAAA TTATGCATTC TTAATTTCTG ATGACAGTAG TCAATACCTG ACTTGGTATA	3840
35	AAAATACGCT TATCGTAGCA TCTGCAAATG CACTGTTTAG TGTGATATTT GTCACGTTAA	3900
	CAGCATATGC TTTTCTAGA TATCGCTTTG TTGGTCGTAA ATACGGGCTG ATTACATTTT	3960
	TGATTTTACA AATGTTCCCT GTATTAATGG CAATGGTCGC AATCTATATT TTGCTAAATA	4020
40	CAATTGGATT ATTAGATTCT TTATTTGGAC TAACACTGGT ATATATTGGT GGATCAATAC	4080
	CGATGAATGC CTTTTTAGTG AAAGGTTACT TCGATACGAT TCCAAAAGAA CTTGATGAAT	4140
	CTGCCAAAAT TGATGGTGCA GGGCATATGC GTATTTTCTT ACAAATTATG CTTCCATTAG	4200
45	CTAAGCCGAT TTTAGCAGTT GTTGCTTTGT TCAATTTTAT GGGGCCATTT ATGGACTTTA	4260
	TATTACCTAA AATACTATTA AGAAGTCCTG AAAAATTCAC ATTAGCAGTT GGATTGTTCA	4320
	ACTTTATTAA TGATAAGTAT GCAAATAATT TCACAGTGTT TGCAGCAGGG GCAATTATGA	4380
50	TTGCAGTACC TATAGCAATC GTATTCTTGT TCTTGCAACG CTATTTAGTA TCAGGTTTAA	4440
	CAACAGGTGC GACAAAAGGT TAGTTTGAAA TTAGGAGTGG GGCAGAATTG ATAAAGAACC	4500

55

EP 0 786 519 A2

	GGGTGTGGTG GTATTGCGAA TGGCAAGCAC ATGCCAAGTT TACAAAAAGT TGAAAATGTT	4620
	GAAATGATCG CATTTTGTGA CGTAGACATT TCGAAAGCAG CGAGTGCGGC AGAAGCATAC	4680
5	GGAAC TGACA ATGCAAAGGT TTATGATGAT TACAAAGCAT TGT TAAAAGA TGACACGATT	4740
	GATGTTATCC ATGTTTGTAC GCCAAATGAC TCGCATTGTG AAATTACTGT AGCAGGGTTG	4800
	CATGCTGGTA AACATGTGAT GTGTGAAAAA CCAATGGCTA AAACGACAGC AGAAGCTCAA	4860
10	AAAATGATAG ATACAGCTAA ATCAACAGGT AAAAAATTAA CAATAGGTTA TCAAAATCGT	4920
	TTCCGAGCAG ATAGTCAATT TTTACATCAA GCAGCGCAAC GTGGCGACTT AGGAGACATT	4980
15	TACTTCGGAA AGGCACATGC CATTTCGTCGT CGAGCAGTAC CAACATGGGG TGTCTTTCTA	5040
	GACGAAGAAG CTCAAGGTGG AGGACCATTA ATCGATATCG GTACACACGC TTTAGATTTA	5100
	ACGTTATGGA TGATGGATAA TTATGAACCA GAATCAGTGA TGGGTTCAAC ATTCCATAAA	5160
20	TTAAATAAAC AGCATCATGC GGCAAACGCT TGGGGTTCAT GGAATCCAGA TGAATTTACA	5220
	GTTGAAGATT CTGCGTTTGG ATTTATTAAA ATGAAGAATG GAGCGACGAT CATTTTAGAA	5280
	TCCGCTTGGG CGATTAAATC TTTAGAAGTG GATGAGGCAA AATGTTCAAT ATCAGGAACT	5340
25	AAAGCAGGTG CTGATATGAA AGATGGTCTA CGTATTCATG GTGAAGACAT GGGTACACTT	5400
	TATACCAAAC ACGTTGAATT GGAAAACAAA GGCGTCGACT TTTATGAAGG TAATGAAGTG	5460
	GATGAAGCTG AAGAAGAAGC AAAAGCTTGG ATTGATGCAG TTGTAAATGA TACTGAACCA	5520
30	GTTGTGAAAC CGGAACAAGC AATGGTAGTT ACAAAAATTC TTGAAGCGAT TTATCAGTCT	5580
	GCAAAATCAG GCAAAGCAAT TTA CTTTGAA TAACATCATA CGGTAAGGAG GCACATCATG	5640
	ACAAAATTAA AAGTTGGTGT GATAGGTGTT GGTGGTATTG CACAAGACCG TCATATTCCA	5700
35	GCATTGCTGA AACTCAAAGA CACAGTCTCA TTAGTTGCAG TACAAGATAT TAATACAGTG	5760
	CAGATGATTG ATGTTGCGAA GCGCTTTAAT ATACCTCATG CAGTTGAGAC ACCTAGCGAG	5820
40	CTGTTTAAAC TTGTTGATGC GGTGGTCATT TGTACACCTA ATAAATTCCA TGCTGATCTT	5880
	TCTATAGAAG CATTGAACCA TGGTGTCCAT GTATTGTGTG AAAAGCCAAT GGCGATGACG	5940
	ACGGAAGAGT GTGATCGCAT GATTGAAGCG GCTAATAAAA ATCACAAATT ATTAAGTGTC	6000
45	GCATATCATT ATCGTCACAC AGATGTGGCA ATTACTGCTA AAAAAGCAAT TGAATCAGGT	6060
	GTGGTTGGTA AACCTTTAGT AGCACGTGTA CAAGCGATGC GTAGGCGTAA AGTGCCTGGC	6120
	TGGGGTGTTT TTACCAATAA AGCGTTGCAA GGTGGCGGTA GTTTAATCGA TTATGGTTGC	6180
50	CACTTGTTAG ACTTATCTTT GTGGCTACTA GGTAAAGATA TGGTGCCGCA TGAAGTGCTA	6240
	GGAAAAACAT ATAATCAATT GAGCAAACAA CCGAATCAAA TTAATGATTG GGAACATTT	6300

55

	GCAAGCATGC AGTTTGAATG TTCGTGGTCT GCAAATATCA AAGAAGATAA GGTTCACGTT	6420
	AGTTTATCAG GAGAAGATGG CGGTATCAAT TTATTTCAT TTGAAATATA TGAGCCCCGC	6480
5	TTTGGAAC TA TTTTGAAG CAAAGCTAAT GTTGAGCATA ACGAAGACAT TGCTGGTGAG	6540
	AGACAGGCGC GTAAC TTTGT CAATGCGTGT TTAGGGATAG AAGAGATTGT GGTGAAACCG	6600
	GAAGAAGCAC GCAATGTAAA TGCCCTTATA GAAGCGATTT ATCGTAGCGA TCTTGATAAC	6660
10	AAGAGCATAC AACTTTAATG ATTATCATAT ATGATACAAA ATTCTCAATA TAAAAAGAAG	6720
	GAGTGCTTTT CAATGAAAAT AGGTGTATTT TCAGTATTAT TTTACGATAA AAATTTTGAA	6780
	GATATGTTAG ATTATGTCTC AGAATCTGGA TTGGATATGA TTGAAGTTGG AACAGGTGGT	6840
15	AACCCAGGAG ATAAATTTTG TAAGTTAGAT GAGTTGTTAG AAAATGAAGA CAAGCGCCAA	6900
	GCATTTATGA AGTCAATCAC AGACAGAGGC TTACAAATAA GTGGTTTCAG TTGTCATAAC	6960
20	AATCCAATTT CTCCAGATCC GATAGAAGCG AAAGAAGCCG ATGAAACGTT ACGTAAAACA	7020
	ATCCGTTTAG CAAATCTATT AGACGTGCCA GTTGTTAATA CATTTTCTGG CATTGCAGGA	7080
	TCAGATGATA CCGCTAAAAA GCCTAATTGG CCTGTTACAC CTTGGCCAAC AGCCTACTCT	7140
25	GAAATTTATG ATTATCAGTG GAATGAAAAG TTGATACCAT ATTGGCAAGA TTTAGCTGAG	7200
	TTTGCAAAAG AGCAAGATGT AAAAATTGCC ATAGAGTTGC ATGCAGGATT TTTAGTGCAT	7260
	ACACCATATA CAATGTTGAA GTTACGTGAG GCTACAAATG AATATATCGG TGCTAACTTA	7320
30	GATCCTAGTC ATCTATGGTG GCAAGGTATT GACCCAATTG CTGCGATTCTG CATATTAGGC	7380
	CAAGCAAATG CAATTCATCA CTTCCATGCT AAAGATACGT ATATTAATCA AGAAAATGTA	7440
	AATATGTATG GTCTAACTGA TATGCAACCA TATGGTAACG TTGCGACAAG AGCATGGACA	7500
35	TTCCGTACAG TTGGTTATGG ACATAGTCCA TATGTATGGG CAGATATCAT AAGTCAACTT	7560
	ATTATTAATG GATATGATTA TGTATTAAGT ATTGAACATG AAGATCCTAT TATGTCAGTA	7620
40	GAAGAAGGTT TCCAAAAAGC TTGTCAAAC TTTGAAATCTG TTAATATTTA CGACAAGCCA	7680
	GCAGACATGT GGTGGGCATA ATACGAACTC GAGGTTAGTC TGAAGTTTGT CTGAAGTAAG	7740
	ACTGGTGGCA GTGTTGAATA AATGCATATG TCGCCAAGCC ATTGCCAAAA ATTTACACCC	7800
45	TTAAATCAAG TCATTGTTTG TAAAGAAGGT GTACTTTATA TAAGTATATA GCGATGGTCA	7860
	TACCCATTCA CAGTAACAAT CCTCACCATT GAAAAGAGTA TATAACCTTT TCAATAGTGA	7920
	GGTATATGAT AATAAAAAA GCCTGTTGTC ACAATGGTCA TAGACACGAC ATACTTTAAA	7980
50	GGTTTCTGAA TATAATATTT CAGAATGCAC TTTAAAGATG GACGTCGATG TAGACTAAAG	8040
	TGATGACAGG CTTTCATCTT TTTAAATATT CATTAATTTT TCTTCTTGTT TAATACGTAC	8100

55

	TAATACACCG ATTAATTCAG GAATGATGTT TAAGAAGTAA TTTGGGTGTT TTGTAATTTT	8220
	ATATAATCCA GATTTAATAA TAGGATGGTT AGGTAAAATG AATAATTTTA ATGTCCAAAT	8280
5	ACCACCTAAA GTTTTAATAA CCATAAATAA CATGATATAA GCAAAGATTA ATATAACTAA	8340
	GCCAATACCA TTTGCAAAGC TAAATGTATC TTTATTAATA AATGCCTCTA CACCAGCCAA	8400
	TACATAAATT AAAACGTGTG TTATTGCTAA AAACCTCGAA TTTTAAACGC CATATTCAAC	8460
10	TGCACCGTCT GCTTTTAATT GTTTTGAGTG ATTAATAGAT ATCTTTAAGC TGACAAGTCT	8520
	GATACAGAAA AAGATAAGTA ATATAGATAG AATCATGATG TCCTCCGTCA TTATGTCATA	8580
15	TGTATAAGCG TTGATTTTGA CAACATAAAG TATTTTATAG ATAAAGCTTG TCAAATACTA	8640
	TTAACTATTT ATTAATTTTA GTACATAAAT ATGTTTCTAA GTATGTGTTT ATGTTTCAGTA	8700
	TTTTGGATAA TTTAATAATT TTAAGGATAT TAAGCGCTTA CACCGACGTG ATATATTTGG	8760
20	CTTAACGAAA ATGATTGAGG TGACAGAGAT GAACTTTTTT GATATCCATA AGATTCCGAA	8820
	CAAAGGCATT CCATTATCGG TACAACGTAA ATTATGGCTT AGAAACTTCA TGCAAGCTTT	8880
	CTTCGTAGTG TTCTTTGTTT ATATGGCTAT GTATTTAATT CGAAACAACT TTAAGGCGGC	8940
25	ACAACCGTTT TTAAGAGAGG AAATTGGATT ATCTACATTA GAACTTGGTT ATATCGGATT	9000
	AGCATTTAGT ATCACGTACG GTTTAGGAAA AACATTACTT GGATATTTTG TCGATGGACG	9060
	TAACACAAAA CGTATTATCT CGTTCTTACT TATCTTATCT GCGATTACAG TTTTAATTAT	9120
30	GGGATTTGTT TTAAGTTACT TTGGTTCTGT AATGGGATTA TTAATTGTAC TTTGGGGACT	9180
	TAACGGGGTG TTCCAATCAG TTGGTGGACC TGCAAGTTAT TCAACGATTT CAAGATGGGC	9240
35	GCCAAGAACG AAACGTGGCC GATACTTAGG ATTCTGGAAT ACATCACATA ATATCGGTGG	9300
	TGCCATAGCA GGTGGTGTG CACTTTGGGG TGCTAATGTA TTCTTCCATG GAAATGTTAT	9360
	AGGGATGTTT ATTTTCCCAT CGGTGATTGC ATTACTTATT GGTATCGCAA CATTATTTAT	9420
40	CGGAAAAGAT GATCCGGAAG AATTAGGATG GAATCGTGCT GAAGAAATTT GGGAGAGCC	9480
	GGTCGATAAA GAAAATATTG ATTCTCAAGG TATGACGAAA TGGGAGATCT TAAAAAATA	9540
	TATCTGGGA AATCCTGTTA TATGGATTCT ATGTGTTTCA AACGTCTTTG TATACATTGT	9600
45	ACGAATCGGT ATTGATAACT GGGCACCGTT ATATGTGTCA GAGCATTTAC ACTTTAGTAA	9660
	AGGCGATGCA GTTAATACGA TATTCTACTT TGAAATTGGT GCATTAGTTG CAAGTTTATT	9720
	ATGGGGCTAC GTATCAGACT TATTAAAAGG TCGTCGTGCA ATTGTAGCTA TTGGCTGTAT	9780
50	GTTTATGATT ACATTTGTTG TCTTATTCTA CACAAATGCT ACAAGTGTCA TGATGGTTAA	9840
	CATTCATTG TTTGCATTAG GTGCGTTAAT CTTTGGTCCG CAATTATTAA TTGGTGTATC	9900

55

	CGCGTATCTA TTCGGTGA CT CAATGGCGAA AGTTGGTTTG GCGGCTATTG CTGATCCAAC	10020
	ACGTAAACGGT TTAAACATCT TTGGATATAC ATTAAGTGGG TGGACAGATG TTTTCATCGT	10080
5	CTTCTATGTT GCATTATTCC TAGGCATGAT TCTATTAGGA ATCGTTGCTT TCTATGAAGA	10140
	AAAGAAAATT AGAAGTTTAA AAATTTAATA TAAATCGGAT TAAAAGTATC GCCAATCTAT	10200
	TGCAATATAG TTGGCAATCC TGCCCCGACG GCATGTGCGT GAAGAGATGA AAGATACTGC	10260
10	TTCTACCCTT GCAAATATAT CATCTCTATG TCTCGGGGCA GATCATAATT CCCTGTTATG	10320
	AAGTATCCTT ATTTGCCCGA CTTAGGGTGA CTCAATGAAT TTAATCCTTA CAATAAAGAC	10380
	ATATAGCGGT GTCAATATTG TAGGGAGTAT TGTTTTATAT TTAAACTCTC TAAAAAGCGG	10440
15	ACTGAAAGAA AAGTGAAAC TTCTCTATCA GTCCGCTTTT TCATAGAACA AAATGGAGGC	10500
	GCCATAATCA TTAGTTATGT GCTAATCTAT TTTGCTTGCT TACAATAATC ACTTGGCGAC	10560
20	ATTTGTAAAT ATTTTTTAAA ATGATAGCTA AACATTTTAT ACTCTGAAAA GCCTACTTTG	10620
	TCTGCAATTT CATAGTGTTT GTAATGTCGA TCTAACAATT GCAGAGATTG TAAAAACGA	10680
	TAGCGATTTA AATAATCGAC AATTGTAATA CCAACATGAT CTTTAAATGT TCGCATCGCA	10740
25	TACGATTCAC TAACATCGAT ATGTTGAATT AAATCTGAAA CAGTCACTTT CGTTTGATAA	10800
	GATTGCTTAA TTTGATCCAC AATCTGGTTT ACATAATAAT CATCGTATTC TACTTTTAAT	10860
	AGTGGTTGGA AGGCATCATG ACAAGATGCT AAGCTACGGC CGTTCTGTGA TTGTTGCTCT	10920
30	AATAAGGTAC GGACAAGTCT TCCTAAAATA ACTTCTAATT GTGCATGGTC TACTGGTTTT	10980
	AATAAATAAT CAAGAACATG ATGTTGAATG CCGGCTTTCA TATATTCAA GTCATCGTAA	11040
35	CTCGATAATA TGATGACATT ACAATCTAGA TGCGCAATAT CATTGAGTAA ATCGACGCCA	11100
	TTTTTACGTG GCATACGAAT ATCAGTAATT ACTAATTCTG GCTGATGTTG TTGAATTAGT	11160
	GATAATGCTT CAACACCATC TTTAGCAGTG TATATTGTAT TGAAATGATA GTCTCCCCAA	11220
40	GGAATGATTT GCTTTAATCC TTCTCGAATA ATTCGTTTCA TCACACAAAT AACTACCTTA	11280
	AACATCTACA TTCCCCCTTG AAAGTGGTAT TTTATAACAA ATTAACGTAC CTTGATTACG	11340
	CTTTGAAAAA ATATGGAGTC GTGCATGTGA ACCATATTGA ATCATTGCTT TATTGTGTAA	11400
45	ATGATTTAAT CCCAAATGCT TAGTATCAAA TACATCATT TTAAGAGATT GGCGTACATA	11460
	TTGCAGGCGA GATGACGACA TCCCGATACC ATTGTCGCAA ACTAAAACAT GTAAATTCTG	11520
	ACGTGCCAAT GTCAGGCGTA TAGTAATGTC CAATGACTCA GTATCTCTAC CATGTTTAAT	11580
50	AGCATTTTCT ATGAGTGGCT GAAGCATCAT TTTACCAATT GTCTGGTGAC GCGCTTCTTC	11640
	AGAACTTTCA ATATGGAGCT TAATCATGTC ATCAAAACGG aTGTTTTGTA TTGCAACATA	11700

55

	GTAACGTAAC ATTTGCGATA ATTGTTGGAC CACAGTTtGT GCTAATTTTCG GAGATAACGT	11820
	AATTAAATAT TGTATTGTTT GCATCGTATT GAATAGGAAA TGAGGCTGGA ATTGGCGTTC	11880
5	TATTTCCCTT AACTGAATAT CACGCAAGCG ACGTTCTGTA TGCTCGATAG AATGGATCAG	11940
	TTGCTCATT GATTCAAATA AATCGTAAAT ATAATTATTA ATTTCTTCTA GTTCACTGTT	12000
	GTTTTTTAAA GGCGTATATG TACCTAGATG ACGATTTTTG GCATAGTAAA TTTTTTGAAT	12060
10	AATCGTTTCG ATATCTTTTG TTTGTCGTTT AGCCATATTA TCTGCGCTAA TGAAACCAAA	12120
	TATTACTAGT AAAACAAGAA CTACGCCAT AACAAATTAAC AACGTGATAC CATCTTCAAT	12180
	GTTTTTCATGT ATATCTTTAT AAATAATGAG ACGATGGTCA GCATGGTTTA ATTTTACAGA	12240
15	TTCAATCATA AATCCGAATT GTTGTGGTcT ATACTTTTCA CCTATAGTAA AACGGTCATC	12300
	GTTGGCGTAT AAAATATTGT CATATTGATC AmCGATAAGT GCGAATTGTC GGTtATCTTT	12360
20	CtTAATTTCA CTAAACGTG GGGTGTtAGC CATATAAAAt TtaAGCATAT ATGTACTATT	12420
	TTTGAATTIA AGCTGATGCG TTGAAAATAA ATACATATTT TTAGTGTTTA AATGTTcATA	12480
	ATTATTGGTT ATAAACTGAT TTGGTCCAGA TAATTCATAA TAAAGTGTTG CGGGCTGTTG	12540
25	GkGTATTAAT TTTAATAATT CACGTTTTGT AGCGGTCACA TCATGATGAT TTGyTAAATC	12600
	GAGCTCTTGA AACGAATTAT TATGCTGTGT AATAAATGTC TGAATCTGCT TTTCAGTATG	12660
	ATGTAAAGAT GACTGACTTT CATCAACATG TTGATGAATC GTACGATGCT CAATCCAAAT	12720
30	ATAGATGGCA TAGAAGCTTA CTAGTCCAAT AATAATGACT AAAAATACTG GAAAAATAGT	12780
	AGACnCAAAT AACGATCGTC TTAATTGATG TCTATAAGGT TTGTATGCCn TCATTGAATC	12840
	ATCTCCAAAA ATTTATGATG TGGAATATCC GGTAATTTAG ATTTCCGTAT TAAAGGTATG	12900
35	TTCTTAAGAT TTTCGATAGA CTGATCGCTT TGTTCACTAA CATCCTTTTCG AATTGACTTG	12960
	GCATcGAACt CTGCAACTAA TCGTtGTTGT ACTGAGCGGC TTGTTAAATA TTGCACTAAC	13020
40	TTTTTACGCT TAGGATGAGG GTGTGCATTT TTAACtAAAG CAATrCCATC AACATTtAAC	13080
	ATTGTTCCCTT CAATTGGATA AACGATTGAT ACAGGATAAC CTTTGTTTTT CCATGTGCGT	13140
	GCATCTTGTT CGTAGCTTAG ACCTGCGTAA TATTTACCTT TTGCAACATC TTCAATGACT	13200
45	TTAGACGTCT TTGACAGTTG CATCGCATGG TTTTGGAATT GATGCACATC ACTTACTCGA	13260
	TGATGCATGC TATAAATAGC ACGCATATGT TGATAGCCTG TCGTTGTTGT ATTTGGATTT	13320
	GAGTACGCAA TTTTACCTTT AAGTATAGGT TGTAATAAAT CTTGATAACC TCGAATCTTA	13380
50	ATATCTCCTT GTAAATCTGA ATTCACtACT ATAACTGTTG GCATTAAATAG AAAACTAGTA	13440
	ACATATTTAT TGTTcGAGCG ATAATCCTCT AATTGCTGTG TTACAGATGT ATCTTGATAG	13500

55

CCACGCTCCG AAAAATCTTC GTTATGCAAG TTTGAAAGCA GTACTTGAGT AGATCCGTGT 13620
 TTAATTTCAA TTTTGACATG CTCTTGTTTT TCAAATTCAT TTAAAATTGG ACGAATCAAG 13680
 5 TTTGATTGAT ACGGAGAATA AACTGTTAAT ACATTTTTAT CGGATTGAGA GTGACGCGTA 13740
 TTAGCGCATG CTGaTAAAAA AATGAGAAAT AATAGCAAGA TATAAATTTT TGATTTTCATG 13800
 10 ATATCCCATC AATTCTATGT ATATTTTAAT ACAATAATTT TAGCAATAAA TGACGCATAA 13860
 GTAATGTTAA ATATTTAGAA ATGTTTATAG ATGACTTGTT AAGACGTTGC AAATGTTGTG 13920
 ATAGCACAAA ATTTTGTGTT GTCAAGACGA TTTACCGAGG CTGTAAAATC AAAGTGTAT 13980
 15 ATTTTATTTG TAGCTGTTAT ATAAAAATCG GCAAGATATT GAACGGTTCA AAAGTGAATT 14040
 TTTACGTCAA TAAAAGTATT TAATCCAGTC TCTTCATATA TAAAAGTAAA TCTTTCTAAG 14100
 TGTGATTTA ACGCTTATCA ACAATCATTT TTTATAAACA AATATATACT CCTAAATTAA 14160
 20 CTTTTAAAGC AATGAAAATA GTGAACATTA TAACTGTTGT GTAACAGAAT GCAATTAGCA 14220
 TATTACTGTT ACACAAATTA GTACAGTTTC TATGTTTTGA CATACATTTG ATGAAAATTG 14280
 TACATAATTT ATGTGAAAAA AATCACAACA AACATGCTAC AATGACTATG AAAACGTTAA 14340
 25 CATAGCATTT CAAATTCACA ACATTATACA GATGGAGGCG TTTAGTATGT TAGAAACAAA 14400
 TAAAAATCAT GCAACAGCTT GGCAAGGATT TAAAAATGGA AGATGGAACA GACACGTAGA 14460
 TGTAAGAGAG TTTATCCAAT TAAACTACAC TCTTTATGAA GGTAATGATT CATTTTTAGC 14520
 30 AGGACCAACA GAAGCAACTT CTAAACTTTG GGAACAAGTA ATGCAGTTAT CGAAAGAAGA 14580
 ACGTGAACGT GGCGGCATGT GGGATATGGA CACGAAAGTA GCTTCAACAA TCACATCTCA 14640
 35 TGATGCTGGT TATTTAGACA AAGATTTAGA AACAAATTGTA GGTGTACAAA CTGAAAAGCC 14700
 ATTCAAACGT TCAATGCAAC CATTCGGTGG TATTCGTATG GCGAAAgcAG CTTGTGAAGC 14760
 TTAAGGTTAC GAATTAGACG AAGAACTGA AAAAATCTTT ACAGATTATC GTAAAACACA 14820
 40 TAACCAAGGT GTATTCGATG CATATTCTAG AGAAATGTTG AACTGCCGTA AAGCAGGTGT 14880
 AATCACTGGT TTACCTGATG CATACGGACG TGGACGTATT ATCGGTGACT ATCGTCGTGT 14940
 AGCTTTATAT GGTGTAGATT TCTTAATGGA AGAAAAAATG CACGACTTCA ACACGATGTC 15000
 45 TACAGAAATG TCAGAAGATG TAATTCGTTT ACGTGaAGAA TTATCAGAAC AATATCGTGC 15060
 ATTAAAAGAA TTAAAAGAAC TTGGACAAAA ATATGGTTTC GATTTAAGCC GTCCAGCAGA 15120
 AAACCTTCAA GAAGCAGTTC AATGGTTATA CTTAGCATAC CTTGCTGCAA TTAAAGAACA 15180
 50 AAACGGTGCA GCAATGAGTT TAGGTCGTAC ATCAACATTC TTAGATATCT ATGCTGAACG 15240
 TGACCTTAAA GCAGGCGTTA TTACTGAAAG CGAAGTTCAA GAAATTATTG ACCACTTCAT 15300

55

AGACCCAACT TGGGTAAGT AATCTATCGG TGGTGTAGGT ATTGACGGAC GTCCACTTGT 15420
 TACGAAAAAC TCATTCCGTT TCTTACACTC ATTAGATAAC TTAGGTCCAG CTCCAGAACC 15480
 5 AACTTAAACA GTATTATGGT CAGTACGTTT ACCTGACAAC TTCAAAACAT ACTGTGCAAA 15540
 AATGAGTATT AAAACAAGTT CTATCCAATA TGAAATGAT GACATTATGC GTGAAAGCTA 15600
 TGGCGATGAC TATGGTATCG CATGTTGTGT ATCAGCGATG ACAATTGGTA AACAAATGCA 15660
 10 ATTCTTCGGT GCACGTGCGA ACTTAGCTAA AACATTACTT TACGCTATCA ATGGTGGTAA 15720
 AGATGAAAAA TCTGGTGAC AAGTTGGTCC AAACCTTCGAA GGTATTAAACA GCGAAGTATT 15780
 AGAATATGAC GAAGTATTCA AGAAATTTGA TCAAATGATG GATTGGCTAG CAGGTGTTTA 15840
 15 CATTAACTCA TTAAATGTTA TTTACTACAT GCACGATAAA TACAGCTATG AACGTATTGA 15900
 AATGGCATT AATGATACAG AAATTGTACG TACAATGGCA ACAGGTATCG CTGGTTTATC 15960
 20 AGTAGCAGCT GACTCATTAT CTGCAATTAA ATATGCACAA GTTAAACCAA TTCGTAACGA 16020
 AGAAGGTCTT GTAGTAGACT TTGAAATCGA AGGCGACTTC CCTAAATACG GTAACAATGA 16080
 CGACCGTGTA GATGATATTG CAGTTGATTT AGTAGAACGC TTCATGACTA AATTACGTAG 16140
 25 TCATAAAACA TATCGTGATT CAGAACATAC AATGAGTGTA TTAACAATTA CTTCAAACGT 16200
 TGTATACGGT AAGAAAACTG GTAACACACC AGACGGACGT AAAGCTGGCG AACCATTGTC 16260
 TCCAGGTGCA AACCCAATGC ATGGCCGTGA CAAAAAGGT GCATTATCTT CATTAAAGTTC 16320
 30 TGTAGCTAAG ATCCCTTACG ATTGCTGTAA AGATGGTATT TCAAATACAT TCAGTATCGT 16380
 ACCAAAATCA TTAGGTAAAG AACCCAGAAG TCAAAACCGT AACTTAACTA GTATGTTAGA 16440
 TGGTTACGCA ATGCAATGTG GTCACCACTT AAATATTAAAC GTATTTAACC GTGAAACATT 16500
 35 AATAGATGCA ATGGAACATC CAGAAGAATA TCCACAGTTA ACAATCCGTG TATCTGGTTA 16560
 CGCTGTTAAC TTCATTAAAT TAACACGTGA ACAACAATTA GATGTAATTT CTCGTACATT 16620
 CCATGAAAGT ATGTAACAAA ATTTAAGGTG GGAGCACTAT GCTTAAGGGA CACTTACATT 16680
 40 CTGTCGAAAG TTTAGGTACT GTCGATGGAC CGGGATTAG ATATATATTA TTTACACAAG 16740
 GATGCTTACT TAGATGCTTG TATTGCCACA ATCCAGATAC TTGGAAAATT AGTGAGCCAT 16800
 CAAGAGAAGT CACAGTTGAT GAAATGGTGA ATGAAATATT ACCATACAAA CCATACTTTG 16860
 ATGCATCGGG TGGCGGTGTA ACAGTCAGTG GTGGCGAACC ATTGTTACAA ATGCCATTCT 16920
 TAGAAAAATT ATTTGCAGAA TAAAAGAAA ATGGTGTGCA CACTTGCTTA GACACATCGG 16980
 50 CTGGATGTGC TAATGATACA AAAGCATTTC AAAGGCATTT TGAAGAATTA CAAAAACATA 17040
 CAGACTTGAT ATTATTAGAT ATAAAACATA TTGATAATGA CAAACATATT AGATTGACAG 17100

55

	TATGGATTCTG ACATGTCCTT GTGCCTGGTT ATTCTGATGA TAAAGACGAT TTAATTAAAC	17220
	TAGGGGAATT TATTAATTCT CTTGATAACG TCGAAAAGTT TGAAATTCTG CCATATCATC	17280
5	AGTTAGGTGT TCATAAGTGG AAAACATTGG GCATTGCATA TGAATTAGAA GATGTCGAAG	17340
	CGCCCGATGA TGAAGCTGTT AAAGCAGCCT ACCGTTATGT TAACTTCAA GGGAAAATTC	17400
	CCGTTGAATT ATAAATACAA TTCAGACCGA AAAGAAAGCA TATGCAACTT CAAGAGTGAA	17460
10	GGGGCATATG CTTCTTTTTC AATTGAGTAT TGAGTATTAG CAAGACGTAG TAAGTATATG	17520
	AGACAACTTC TACAATGGTT GAAGGAAGAC GTTTTTGTAA GTAGCTATGC TGATAAAGAA	17580
	TGTGATGTCT TGTAAAGGT GGGGTTCCAA TATCATCATT TAGCTGATGT TGAATGGGTT	17640
15	ATTATTTGCT ACTTGCATAT GAATATGAGT CTTTTCAAAT TTTTATTGAC CCTGAGTAAT	17700
	GAAAAATATT AAGATGAAAC TTAATATTAA AgCAATGCGG AGCGTGATTA TGAAGAGAAT	17760
20	TAGTAAAGAT ATATGGGCAG TATTTAAATT ACTGTATCaA AATAAAGGGC GTTTTAGCAT	17820
	TAATGCCTTA CTATTGCAGT TAATCATGAT TTTTATTAGT AGTACATACT TAATTTTACT	17880
	ATTTAATATG ATGTTAAAAG TAGCTGGcAA AGCCAACTTA CGATTAAcAA TTGGACGGAA	17940
25	ATCGTTAGTC ATCCCGCCAG TGTGATACTT CTTATTATAT TCATATTAAG TGTTCCTTT	18000
	CTGATTTATG TAGAGTTTTT ATTGTTAGTT TATATGGTTT ATGCCGGCTT TGATCGACAG	18060
	ATTATTACAT TTAAATCCAT TTTTAAAAAT GCCTTTGTAA ATGTGCGTAA ACTCATAGGT	18120
30	GTACCAGTTA TTTTCTTTGT CATTTATTTA ATGTTAATGA TACCCATTGC CAACCTAGGA	18180
	CTAAGTTCAG TATTAACAA AAATATTTAC ATACCTAAAT TTTTAACGGA AGAACTTATG	18240
	AAAACGACGA AAGGTATAAT CATTTACGGT ACCTTTATGA TTGCTGTATT TATATTAAAT	18300
35	TTTAAATTAA TATTTACTCT ACCGTTAAG ATTTTAAACC GCCAGTCGTT ATTTAAAAAT	18360
	ATGAGACTAA GTTGGCAAAT TACGAAGCGA AATAAGTTTC GGCTTGTTAT AGAAATAGTT	18420
40	ATATTAGAAC TCATCATTGG TGCGATTTTA ACATTAAFTA TTTCAGGAGC AACATATCTT	18480
	GCTATTTGTG TAGATGAAGA AGGAGATAAG TTTTGTAGTCT CATCAATTTT ATTTGTTGTA	18540
	TTGAAAAGCG CATTGTTCTT CTATTATkTA TTtACGAAAT TATCATTAAAT CAGTGTGTTA	18600
45	GTACTGCACT TAA	18613

(2) INFORMATION FOR SEQ ID NO: 113:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1214 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

	AAAGTTTAA AAGGGGTGAG ATACTTGGCG AATAATCCAT TCCAGCTTTG CGTTTAAAAG	60
5	GAAITATACT TGCCATTGTC GGTGCTTGTT TATGGGGATT AGGTGGTACT GTTCTGATT	120
	TCTTGTTCAA ATATAAGAAT ATTAATGTCG ATTGGTACGT CACTGCTCGA CTTGTAGTCA	180
	GTGGTGTTTT CTTACTTATT ATGTACAAAA TGATGCAACC CAAACGTTCA ATATTTAGCG	240
10	TATTCCAAGA TCGACGTATG TTAGGCAAAT TACTTATCTT CAGTATACTG GGCATGTTAG	300
	TAGTACAATA TGCTTATATG GCATCTATTA ATACAGGTAA TGCTGCGATT GCAACATTAC	360
15	TACAATACAT TGCGCCAGTT TATATTATTA TTTGGTTTGT CATAAGAGGC GTTGCAAAAC	420
	TAACATTATT TGATGTGCTT GCTATTATCA TGACACTATT AGGAACATTT TTATTATTAA	480
	CAATGGTTC ATTTTCTAAT TTAGTCGTCA ATCCTGCAAG TTTATTCTGG GGTATTTTAG	540
20	CTGGTGTAGC ACTCGCTTTT TACACAATTT ATCCTTCAGA CCTACTTAAC CGCTTCGGTT	600
	CGATTCTAAT TGTCGGGTGG GCAATGCTTA TTTCTGGTGT TGCGATGAAT TTACGCCATC	660
	CAATTGGCA CATTGATATC ACTAAATGGG ACATATCAAT TATATTATTT TTAATCTTTG	720
25	GTATTATCGG TGGTACCGCA CTCGCATTTT ATTTCTTTAT CGACAGTTTA CAATACATAT	780
	CAGCGAAAGA AACCAACATTA TTCGGAAGTGT TTGAACCTGT CGTAGCCGTT ATCGCAAGCA	840
	GTCTATGGTT ACATGTGGCA TTCAAACCAT TTCAAATCGT AGGCATCATT CTTATTATGA	900
30	TTTTAATTTT ATTACTATCA CTAAAAGAC AACCTGAAAC ATTAGATGAA TAAGAAACT	960
	CTGATAATCA CTTTAGCAAG TAACTATTAT TTAACAACGT AGTTACCTTA TAGGTGATAT	1020
	CAGAGTTTTT TATTTTAGTT AATAATATTT TTTCACTGGT ATAAAAAGC GTCGTCGCTC	1080
35	TGGTAATCGG AAATACTGGA ATAAAATATG GAATTGGGTA ATAATCCCAG GTAnTAAAG	1140
	TCCATGTTCC GATAnCCTnT CCGCAnCTCC AACCAAATTT GCCGATAAGG TTCCAAAAGG	1200
40	CATCCTGGGG GTAC	1214

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

- | | |
|----|-----------------------------|
| 45 | (A) LENGTH: 9458 base pairs |
| | (B) TYPE: nucleic acid |
| | (C) STRANDEDNESS: double |
| | (D) TOPOLOGY: linear |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

50	ATTTTGTTTT CATTCACGAT GGGGTnATAC AGCAAACACA nCTAAAATAA CTATCAATAG	60
55		

	CTTAGACAAT AAAAAATATG CCACTACAAT CGCTAATATT ACGATTAAAA AAGAAGCGTT	180
	AACGATTACT TTCATCGTTG TTCTATCTCT GAACATCATA TTAAAGACAA CTAGACTAAT	240
5	TGATAATGAA ACAGCAAAAA AAGTAATAGC TAACACTAAT TTCATCATAA ATAGACAGAC	300
	TAAACCTATG ACTAATAATG TATTAGAAAT TACAGCTGAC GTTTTAAACA TTCTCGaATT	360
	AATATGCACT CACCCTTTTT ATTTAAATAA CTTACATAAT CATAATAATA CATGATGTTT	420
10	CATAGGCCTG TCGATGATTG ATTCACAATA GCACGTGATT TTTTGTGTTT TCAATATTAT	480
	TCATTTATTC CATCAAAAAC ACCCTTTTTA ATTTTACAA AAATTAAAAA AAGTGCTCCT	540
	ACACTGCTTG CATGTAGAAA CACTTTTTCA TTGTAATGTT ATTCCTCTCG AGACATACCT	600
15	TTTAGCATAT TAAGCATGTA TGTAAACTA CGGTTTCATGT CGTCATCTTT CAATACGCCC	660
	AATAGACTTC TTATAGTTGT CTTAGCATTG GGACTCGCTT GATTGGCAAC GTGTAATCCT	720
20	TTATTAACCTT TATTAGGAA GTCGCTTAAA TCTGATACAT TGAGTTCACC TAATAAAAAT	780
	ACCATGAAG CCATATTAGA TAATAGCCCT GTATAAATAT CTTTATTAAG TTCAACTGCA	840
	AAITTTATTTA TGATGACTTG ACGTCCTCGA ATTGCACCAT TTAAAGCATC TAATAGTTTT	900
25	GCATCATCTA ATGTTTTAAT AAGCTTGATT GCTTTAATA TACTATCTTT ATTCGCTGCA	960
	ATTGCCTCTG TAACCTCATT TAACTTTCT AACTTAATTT GTTCTTCTGA TTTTCTAAG	1020
	CGTCTAATTT TAGAAGATAT TCTCTCAGCC ATTATTTATC CACCTGATTT CCCGGGAAAA	1080
30	CATAATCTGA ACGTTCCCAT TTTTCTGTGA CTTGAACACT GTACTGCGGT TGACGTTTTT	1140
	TATTGACACG GAAATTATTA GGGTTCAACG GTGACTTACC ACGTTTCGTA ATTACCTCCA	1200
	AACGACAGCT AGTACGTTTA TAAGATGGTG TATCCGTGTA TTGATCAACA TCACTaTTAG	1260
35	TTAATAAGTT AATTGCACCT AGATCTCCAT TTTCCATCGC aTCaTTATTT AATGGAATAT	1320
	AGAItTCTTT ACCTTTAACA CGATCTGTCA CGTGAACCTG TAATACCGCT TCTCCTGTYT	1380
40	CAGAAATCAG CTTAACTTCT GCACCTTCAT GAATGCCTCT ATCTTCAGCA AGCTCTGGAG	1440
	AAATTTCAAC AAATGCACGT GGCACCTTGT ATTTAATCAT TGGTGTTTGA TAAGTCATAT	1500
	TACCTTCATG GAAGTGCTCT AACAAATCGAC CATTGTTTAC ATGAATATCA TAAATTTTCA	1560
45	CTTGCTTAAA GTAATTATCA AATGATAATG GGAATAATTT TGCTTTACCA TTATCAAAAT	1620
	TGAATCCTTC TAAGTATAGA ATAGGCTCAT CAGTACCATC AGGTTGTAAT GGCCATTGTA	1680
	AACTATTGAA TCCTTCTAAA CGATCATAAC TTACCCAGC ATATAGAGGT GTTAAGCGTG	1740
50	CTACTTCATC CATAATTTCA CTAGGATGCT TGTAATTCCA ATCAAATCCT AATCTATTAG	1800
	CAATTGCTTG GAAAATTTTC CAGTCAGGTT TTKAATCACC AAGAGGTTCT AATGCTTGGT	1860

55

	TTGCTGGCAA TACAACATCT GCGTATGTTG CTGTGAATGT TAAAAATTCA TCTTGGA	1980
	CCATGAAATC TAATTTTCA AACGCAGCTT GTACAAAATT AATATTTGAA TCCACAATAC	2040
5	CCGTATCTTC ACCATATAAG TACAATGAGT GTACTTCTCC GTCATGTATA CCTTCTACCA	2100
	TTTCATGATT ATCTTTACCA GCTTTTGGAT TCAATTTAAC GCCATATTCT TTTTCAAATT	2160
10	TAGCGCGAAT ATCATCCGCT TCAATACTTT GATAACCAGT AATCTTATCA GGCATACTTC	2220
	CCATATCACT ACATCCTTGA ACATTATTAT GTCCACGTAA TGGATACGCA CCAGTACCAG	2280
	GACGACGATA ATTACCTGTT ACTAATAATA AGTTTGAAAT CGCTGTACTT GAGTCACTAC	2340
15	CAATGTCTTG TTGTGTAATA CCCATTGCCC AACAAATTAC AACAGATTCA GCTTTAGCAC	2400
	ATTCTTCAGC AAATTTAATC AATTCTGATT CAGGAATACC TGTTGCTTCT TCAGCAAAAG	2460
	CCATTGTAAA TGTTTCTAAT GATTTGTAAT ATTCATCAAA ATCATCTACC CACTCATCAA	2520
20	TAAATGCTTT ATCGTGTAAT TCATGATCAA TAATATACTT AGTCACTGCA CTTAACCACG	2580
	CTAAATCCGT ACCTGGTTTA GGTGATAAA AACGATCCGC ACGTTCTGCC ATTTCATGTT	2640
	TTCTAATATC AAATACATGT ATTTTTTGAC CAAATAATT TTGTGCACGT TTCATGCGTG	2700
25	ATGCGATAAC TGGATGAGCT TCGGCTGTAT TAGTACCTAT CAATACAGAC ATTGCCGCTT	2760
	TTTCTAAATC TTCAATACTA CCTGAGTCAC CGCCGTGTCC AACCGTTCTA AATAAGCCTT	2820
	TTGTTGCAGG TGCTTGGCAA TATCTTGAAC AGTTATCAAC GTTATTTGTG CCAATAACTT	2880
30	GTCTTGCTAA TTTTGCATT AAATACGATT CTTCAATCGT CGCTTTAGAA GAAGAAATGA	2940
	ATGATAGTGC ATCTGGGCCA TGCTTTTCTT TAATAGCTGT AAAATTATCT GCAATGACGT	3000
35	TTAAAGCTTC ATCCCATTCT ACTTCATGGA ACTCACCATT TTTCTTACT AGTGGTTTAG	3060
	TTAATCGTTG ATCTGAATTA ATATGTCCCC ATGAAACTT ACCTTTAACA CAAGTCGCAA	3120
	TTTTATTGTC TGGAGAATCA TGTGATGGTT GTACTTTTAA AATTTCTCTA TCTTTAGTCC	3180
40	AAACTTCAAA TGAACAACCC ACACCACAAT AAGTACACAC TGTTTTAGTT TTCTTAATAC	3240
	GCTCTTTACG CATTTCTGCT TCTGAATCTG AGATTGCAAA TAGTGGACCA TAACCAGGTT	3300
	CTGCTTTTTT AGTTAAATCA ATCATTGCTG CTAATGAACC AGGTTCCGTA TCAGTCATAT	3360
45	AACCCGCATT ACCTTCCATA TTCACTTCCA TCATGGCATT ACATGGACAT ACCGTCGCAC	3420
	ATTGACCACA AGATACACAT GAAGACTCAT TAATCGGTAC ATCATTATCC CAAATAACAC	3480
	GTGGATGTTT ACGATCCCAA TCAATTCTAA TAGTTTCATT CACTTCGATA TCTTGACATG	3540
50	CTTCTACACA ACGCCACAT AAGATACATT GATTTGGATC ATAACGATAA AATGGGCCGT	3600
	AATCTTTTTT GTATGGCTTC TCTTTATATT CATACGTTTG ATGCTGAAGC CCCCATGCAT	3660

55

	TATGCTTTTC TAAAATTGGA TCAAGCGCTT CTTTTTGAGC ATCTTTCACA TCATTGTTCA	3780
	CAGTATTTAC AGTCATTGGA CGATCAATCA CCGTACTACA TGAACGTTCA ATTTTACCGT	3840
5	CAATCTCAAC AGTACATGTA TCACATGTTT GAATTGGTCC CATCGACTCG TTATAACAAA	3900
	TTGAAGGTAC AAAAGTATCT TGTGATTAA TAAATTCAAG TAAATTCGTA CCTGGTTCTA	3960
10	CAAGATAATC TTTTCCATCA AGTGTAACCA CCAAATGTTT TTGCATATTA CTCACCCCGT	4020
	CTATATATAT TTTCCGTAAA TGACTTTTAA TAAATTGCTC ATATCCACCT AAAATAACGA	4080
	TGCCCCACAC ATCTTTCAGA TAGAATTAAT TTAATTGTAT TACTTTATGT ACTAGTTGTT	4140
15	AAGTAAATTT TTGTATTTTG CCTTTTTTACA ATCATTTTTA TTTGAAATAT TTTGCGCGAA	4200
	ATTAAATCAT CTTTTTGTTT AATTGAAAAAT AATTATCATT ATTAGTTTTT CAATTATCTG	4260
	TTTCACGCTT TTTGCCATAT CTTTCACAAC CTTATTAATG ACAATATTTA ATAATCACCT	4320
20	CACCTAAAAA TCGTTATACT ATTTATAAAT ACCCTTTTTT TGAAAATTAA TAACCCAAGT	4380
	TTGATAAATA TCTACTATCA TTTAGAAGGT AATATTTATC TTTAAATTAA ATTTGTAATG	4440
	GATTAATTTA TAAAAATCAA ATCAGGCATT AAATAAAATA GCCCATAAAT ACAAAGTGTT	4500
25	ATCACCTTCT ATTTACGGGC TATTAGTTCT ATTCGTTATT CTATTTACAG ATCATTCTAT	4560
	CTAATTAATT TGTGTACAAT TTTGATAACT TATTTTCCCT TAGTTTACTA CTCTAGATTA	4620
	TCTTTTAATA ACTTAGTACT TTCAGCTTTT GACTGCTCAC TAGGAATGAA GTAGTACAAT	4680
30	CCGTCACTTT GAATGCCGCC TTGACCACTC AATTGATGTT TATTAATCGT GTCATTAGCA	4740
	TCTTTATAAT TGCTTCTAAT CGTATTCAAA TCACCTAATG TTAAATCTGT TTTAACATTA	4800
35	TTTTGAATTT CATTCAATAG ACTATTAAAA TGTGTAATCG ATGATGGGCT TGCAATCTTA	4860
	TTGGCCATCG CTTCAAGCAC AATTTGCTGA CGTTGTTGTC GACCAAAGTC ACCACCAGCA	4920
	CCTTCTTCTT TACGACTTCT AATAAACTTC AATGCTTGAT CACCATTAC ATGTGTCTGC	4980
40	TGTCCTTTTG TAAAACGAAC ACCATCAACA GTGAATGTAT CATTACTTAC TACATCAACA	5040
	CCGCCGATGC TATCTATCAT ATTATGCAAA CCATCCATAT CGATTGTCGC ATAATGATCA	5100
	ATTGGCACAT TCATTAATTT TTCAAGTGAT TTAACAGCCA TATTTGGTCC ACCATATGCA	5160
45	TAGGCATGTG CAATTTTTTC AGTAGTACCA CGGCCAACAA TTTCCGCTCT TGTATCACGC	5220
	GGTATACTTA CTATTTCACT TTTCTTCGTT TTAGGGTTGA TAGATAAAAT CATAATACTA	5280
	TCCTACGCT CTCCGCCACC CTTTTTCTTA CGATCAGCAT CTGAATCGAC ACCAAATAAA	5340
50	GCGATTGTGA ATGGATCACC ATCGTTTAAA CTCACTTTTT TATCTCTTAA TTCTGAATGA	5400
	TTGCGATCTA ACGGATTGTG TATCTTATTA CCAGTAATAA AAATTTTAGC AGCTACATAC	5460

55

	GGTAGGCTCA TTTTACTTTT AGACGAACGT TTCAATCCCA CCACTCCTTT ACTATTCCTT	5580
	ACATACTTTG TCTGTTTTCT CTATTTATTA TATAGTAAAA TAATTTTTTTT ACTATACTTC	5640
5	TGTAGACGTA TAACTATTTT TTATCATTTT TTATCTCTAG AGAATATCTA TCTGTATTTT	5700
	TGATAACCAC CATTTCGATT TAAAATTTTA AGTACCGTTT CATGACATGC TTTATTACTT	5760
	ATAATAAAAG GTGCACCCTT TAAATGATCA ATTGCCTTAC CATCTAAAGT CGTCATTTTTT	5820
10	AGATTCAATA GTTCTGCAAA TAAAACTGT GCAGCAATGT CCCAAGGTTT AGGATTTGTA	5880
	TTAATATGTG CCCCAAATTG ACCTTTTGCC ACTCGCATAG AATCTAATCC GCAAGCACCA	5940
	ACTAAACGAT AACTAAATGA GCGTCAAAT AAATCTTGCA CCGTATCTAG ATTCATCACT	6000
15	TGTGCATTAA ACGATATAAT AGCGTCTTCC AATTTTAACG ATGGTGGTTC TTCCATCTTA	6060
	ATTCCATTAC AAAAAGCACC TTCTCCTCGT ATTGCTTTAT AAAGCTTTTTT ATGCGGATAA	6120
20	TCATATACGT ACGATAACAT TGGTTTACCT TCATAAAAT ACGCCAATAT AATACAATAA	6180
	TCTTCTTGCT GTTTTACTAA ATTGGCAGTT CCATCAATGG GATCCATAAT CCATAAATGA	6240
	TTAATTTTAT TCGTAATCAT TTCATTACTT TTTTCTTCCG CTAATAGTTG GTGTTCCGGA	6300
25	AAATGTGTTG CTAAAAATTG TTGGAATTGT TGTTGAATCT GTTTATCTAC ATTTGTAECT	6360
	AAATCAAATC GATGACGCTT AGTTTCTGTA GTCATTTCCA TAATTAATTG CGGAATAACA	6420
	TTGTCTATTT GTTTCACCA CGAACATATT AACTTATCTA TTTGCTGTAA TGTTTTATCT	6480
30	GTCATTTCTG CCACCACTTC TCATATCATT ATCATTTTAT TATTACCCTA TATTAAAAGA	6540
	ATCAACAATA CAACTGAAGA CTTCTTCATT TTATGCATAA AAAAATCGGC TAGTCACGTG	6600
	CTAGCCGACA AATAGAAAGG AAAGTAAGTA ATAAATATTG AAGATGTTGT GATGTAACCTT	6660
35	GAACGATTAA AAGCTATCTG TTATATAGCT CTACCCCTTT GTTTAATCGC TCCCCCTGTT	6720
	ACAAGTAATA TCATAGCACA ATCTTTTTTA AAATGTAAGC GTTTTCCACA AAATTTTTAC	6780
40	GATTTTTTTA AAAAGATATT GAAAATGTCC TCATTGTCAC TCTTATGTTA TACTTTGTGT	6840
	AATATATCAT CTTTTAGGAG GTGGCTGTCA TGAATAAAGC TGAAAGGCAA AATTTAATAA	6900
	TTACTGCAAT TCAACAAAAT AAAAAAATGA CCGCTTTAGA ATTAGCTAAA TATTGCAACG	6960
45	TATCCAAACG CACAATTTTA AGAGATATTG ATGATTTAGA AAATCAAGGT GTTAAAATTT	7020
	ATGCGCATTG TGGGAAAAAT GGTGGTTACC AAATACAACA AGCACAACTT AAAATTGCAT	7080
	TAACTTATC TGAAACACAA TTATCAGCCT TATTTTTAGT GCTTAATGAA AGTCAGTCGT	7140
50	ACTCGACATT ACCATATAAA AGCGAAATCA ACGCAATTAT AAAACAATGT TTAAGTCTTC	7200
	CACAAACACG CTTAAGAAAA TTGCTTAAAC GCATGGACTT TTATATTAAA TTTGATGACA	7260

55

	ATGTGATGTT AGTAGATCAT AGGGTTGATG ATAATATTAA AGCTGAAAAC GTTATATTTA	7380
	TTGGCCTTTT GTGTAAACAT GGACATTGGC ATGCAGTCAT TTATGACATT GCTCAAGACA	7440
5	AAACTGCCGA ACTCGAAATT GAAAATATTA TAGATATTTT GTATTCATTG GGTAAGACGA	7500
	TTCAAACCAG AGACATATCC ATTGATAACT ATCATCAATT TTTAAACCCC ATCGATTCCCT	7560
	AAAAACAGC AGTAAGATGA TTTTCAATTA GAAAATATCT TGCTGCTGTT CTCTATTTAT	7620
10	ACAATACTTC GTATTGAATG GntTCGCTTT CCTAGGGTGC CGTCTCAGCC TTGGTCTTCG	7680
	ACTGGCACTG CTCCCTCAGG AGTCTCGCCA TTAATACTAC GTATTAACAT GTAATTTTAC	7740
	TTTGAATAC TTAaaaaaat AAAACACTTT GCCCAACTTA CACTACCAAT AGAAACTGCT	7800
15	GTTAGAATTC CTCAAATGA TATTTGCGCA TATGTTAATG AAATTGTTAA AAAGATAGCT	7860
	GATAGCGAAT TCGATGAATT CAGACATCAT CGTGGCGCAA CATCCTATCA TCTAAAAATG	7920
20	ATGTTAAAAA TCACCTCATA TTCATATACT CAATCTGAAT TTTCTGGCCG TAGAATAGAA	7980
	AAATTACTTC ATAACAGTAT TCGAATGATG TGGTTAGCTC AAGATCAAAC ACCTTCTTAT	8040
	AAAACATTA ATCTTTTtag AGTGAATCCT AATACTGATG CGCTAATTGA ATCTTTATTT	8100
25	ATTCAGTTTC ATAATAAAAT GCATATCAAA AAAGCTGATT TCTATCAAAT AATTAATAGA	8160
	AATCAGCTTT TTTCaTTGCC TAAAAACTTA ATGTCCCGAC CTCTTTATCT ACGCATAAAT	8220
	ACTTATTACT GATATAACGA AAGAAACAAA ATTATTTGCT ATATGTAATG CAATTGTTGA	8280
30	ACCTAGGTTT CTTCAGATT TTAAATAAGT GAAACTAAT ATGATGGATA GTATGAGATA	8340
	TGGACCAAC TCAAACGGCG ACTTTGCATC AGTCACATGA ATAAATGCAA ATAAGAACAC	8400
	CGAAACAATA CTCATAGCTA TAAATTTAA CTTCCTACCT AATTCTCCAA TTAAATATG	8460
35	TCTAAATACG ATTTCTTCAA CTATTGGACC TACAATCACA ATTAATAAGA ATGCTACAGG	8520
	TAAAAATGCA GGCACCTTCAA ACATTTTATT TAGCTCAAGT TCATTGGCTG TTtCACTATA	8580
40	TTGCAAATGT TTAGGTAGAA ACTGTGTCAT ATATTCATAT GTATAAATTA AGATGAGAGC	8640
	AATAATATAC GTTATTGACA ATCTAAGCCA ATATTTTTTG ATATACGCAA AACCAGCTCG	8700
	AAGCCTTGAT GGCATCACTT TTAAATGAAA TAAATAAAAT GCGCCAATCC CAATCGTATA	8760
45	TGCTAAAGCT TGTGTGATAG TCGCTACAAA TATCAGATTA CTATCGATT TATAATAACC	8820
	AAACAAAATT GGTCTATGT AAGCTGCAAT TGTGAGTGCA TAAATATAA CACCTATAAT	8880
	TGGAATTATA AGCAAATCTC TCCATGCTAT ATCTTTAAAC GTGTATTTCT TTTTTTCATT	8940
50	TTCCaCTGTT ATATCctTTC CTGTTTAATA ATTGATTTTT GGAGGTACTT CTACATGATA	9000
	AACGAAACTA AGTATATGAG ACAACAAATT ACTAATTTGA TTCAAATCAT TGATACGATT	9060

55

ATAGTTACTA ATGAATTGAA TAAGTTCAAA GGCTTTGAAA CATCATATAT AATAAACGAA 9180
 AATCAAGTTT CCTATTATGA AATTATAACA CTACTTAATA AACGTCCCCT CgACAAGTCG 9240
 5 ACTATGGTAA CAAAATTCAA TATCTTAATT TTTATCATAC AGAACTATCT AACGCATTAT 9300
 TTGCAATTAA ATTTGCCCAT TAACCTATTT TTCATAAAAT GTCATTTAAA CAAGTTATTT 9360
 ATTAAATTC ACTTTATTAC ATAAATTATA CAATTArAAA GTTTCCTCAA ATTGTAAAGA 9420
 10 TGCATTAATC GAGTTATAAT CATAATGATT AAGATGGT 9458

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 910 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

AnGCGTATCA TGTCACGCAT TTTAACTACT TCTTTACCAC AAGATTATAC AGTCACATTA 60
 25 GTTGATCGTA TGCCATTTC A TGGATTGAAA CCAGAATTTT ATGCTTTAGC TGCGGGCAGC 120
 AAATCAGATA AAGATGTTCTG TATGAAATTC CCTAATCATC CACAAGTGAA TACAGTTTAT 180
 30 GGTGAAATTA ACGACATAGA TTTAGATGCT CAAATTGTCT CAGTCGGTAA TTCTAAAATT 240
 GATTATGATG AGCTAATCAT TGGTTTAGGA TGTGAAGATA AATATCATAA CGTTCCAGGA 300
 GCCGAAGAAT ATACACATAG TATTCAAACA CTCTCAAAGG CTCGGGATAC TTTCCATAGT 360
 35 ATTAGTGAAC TACCAGAAGG TGCTAAAGTC GGTATCGTTG GTGCTGGATT AAGCGGCATA 420
 GAACTTGCCA GCGAATTAAG AGAAAGTAGA TCAGACTTGG AAATATATCT TTATGACCGT 480
 GGGC[~]CGCGAA TTTTAAGAAA TTTTCCAGAA AAATTAAGTA AGTATGTTGC GAAATGGTTC 540
 40 GCCAAAAATA ATGTTACCGT TGTTCCAAAT TCAAATATTA ATAAAGTTGA ACCTGGTAAA 600
 ATATATAACT GTGATGAACC TAAAGATATT GATTTAGTTG TATGGACAGC AGGAATTCAA 660
 CCTGTTGAAG TTGTTCTGTA CTTGCCGATT GATATAAATA GTAATGGACG CGTGATAGTT 720
 45 AACCAGTATC ATCAAGTACC AACATATCGT AACGTCTATG TAGTTGGTGA TTGTGCTGAT 780
 TTACCACATG CGCCAAGTGC TCAGTTAGCC GAAGTTCAAG GTGATCAAAT TGCCGATGTG 840
 50 CTTAAAAAGC AATGGCTAAA TGAACCATTA CCTGACAAAA TGCCGGAAC TAAAGGTACAA 900
 GGTATCGTTG 910

(2) INFORMATION FOR SEQ ID NO: 116:

(A) LENGTH: 10182 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

10	TTTTTGATTG AAAGTGGTGA TTTAACAAGC ATTTTAAATA GCAATGATTG GAAAGTCACA	60
	CATGATCCTA CCACTGATTA TTATAATTTA TCTGGTAAGT TGTCGAACGA TAATCCAAAC	120
	GTTAAACAAT TAAACGTAG ATATAATATT CCTAAAAACG CATCAACAAA GGTGGAATTA	180
15	AAGGGAATGA GTGATTTAAA AGGCAATAAT CATCAAGATC AGAAACTTTA TTTTATTTT	240
	TCAAGTCCTG GAAAAGACCA AATCATTTAT AAAGAAAGCC TTACTTATAA TAAAATAAGT	300
	GAACATTAAT ACTTATGCTG TAATTATAGA AACATCCAAA TCATCTATTA TAATCCTATA	360
20	TTATAAAAnC ACCTCACATA ACTCGTTCAA CTGTACCAAA CCACATTACA TTAGATTTTA	420
	GGCTAACTAT TGTGATGTAC ATCAAAAACG AATTTGTGAG GCGTTGTATA TTTTACAAAG	480
	GTGACTAGCG TTTCTGATAG CATTTCCAAC ATTACTACAC TCAAGCGTCA CGCTAAAGTT	540
25	CGAAATCGAA TCCTTTCATT CAACAAAAGC TCATATCCAC TACAACTTC ATATCAAGCG	600
	TATAAACTAT CTTGTGATAC TATCTCGATC ATATCTATAG TATGCATTG TGTTCCGTTT	660
30	CACTGAAGTA TATGTATCAT CAGTTAAGTA TAAACCGTCA TCCTTCAATG TTAATTGATA	720
	AGCATATTTT CGTGCTAACC AGGCAATATC TATATAATTT TCTCCTGCGT TTTCATAACT	780
	TCTTAAATCT TCAATATGTG CACTAACTTC AGGGaAAATG ATTCTAACAA CACTTTCATC	840
35	AACCCAATAT TTGTCATGCA TCCATCGCAC TTGATCTGCC AATAAAGGTA ACTGCACATC	900
	ATTGAAATAT AGACGAAAGC CGTCACTATC ATACATTTGC CGATATGGTA ATGGCTGTTT	960
	TCTAATCACT AACACCTCGC CACCCATTAC GGTGCCTTCT CTAGTATCAT CACTTCCACC	1020
40	CGAAGCTTCA TACGTTGTTG GGTCAACCTG TAGTCCATGT ACATCTCCAA TATAAGCATC	1080
	TGGTTTATGT TCCATTGCAT GTCCATGTGC AATCAATGCT AATATTGTAG ATTGTGAAAA	1140
	TTGAGGCTCC CATTCAATGC GATTAGGATG GCTACTATAA ATTCTAGGTT CATCTATAGC	1200
45	CTGCTGAATA TCCATGCCAA AACTAATAC ATTGATTAAT GTTTGCGCAA CACTAGCAAT	1260
	GATACTTATG GCACCAGGTG CACCTACTGT TAATATTGGC TTCCCGTGAT ACATCACAAT	1320
50	CGTTGGAGCC ATGTTACTTA GTGGTCGTTT ATATGGTGCA ATTTCTGTAA TACCACCATC	1380
	TACTACATCA AAGCCATCCA TTGTCGTATT CAATAACACA CCGTAGCCTG GAATCGTGAT	1440
	ACCTGAACCA TAAATCATAC CAATTGATGT CGTAAATGAA GCAATATTAC CTTCTTATC	1500

55

	ATCAGACACA ACACCATGCT CTATATCAAT ATTTGCTTTA TTGCTATCAA TGAGCGTACT	1620
	GCGTGCTTTT AAATAATCAT CATCAATTAA TGA CTGTACA GGCACCTCAT GAAAATTATC	1680
5	ATCCGCCAAG TATTGCGCAC GATCACTATA TGCTAAATGC ATCGCTTGTA TCAAATGATG	1740
	CAAGTAATCA ACAGATCTTG GACCCATAGA TGGTAAATCG ACATGTTCTA ATAACTTCAA	1800
10	TATTTGAATT ACCGTGATAC CGCCAGAACT AGATGGTCCC ATTGaATAAA TGT CATAGTC	1860
	TTTAAATGTT GCACTGATTG GCGCTTTAAT CTGAATGTCA TATTTGGCTA GATCCTCTAA	1920
	AGTGATTGTC CCACCACATG CTTTGACAAC ATTGACTAAT TGTTTCGCAA TGT CACCTTT	1980
15	ATAAAATGCA TTAAACCCTT GTTCTCTTAA TATTTGAAAT GTCTTACCTA ATTCGGGTG	2040
	TACAATCCAA TCACCTTCAC GCCAATATTG ATTTTCATGC GTAAATACTT GTGCCGTTTC	2100
	ATGATACTTT GTCAATCGTG CGTGTGCTG GCGCGAATAT TTTTCAGTAG CCCAATTGGC	2160
20	TGCATGACCT TCAATGGCTA GTTCAATTGC AGGATTAAAT AAATCTTCCA ATGACAATTT	2220
	AGCATAACGC TTGTGAATAT AATCAAACAG CTTTGGAATT GCTGGCACAG CGACAGTTTT	2280
	ACCATGTGTA GTCATATCAA AAAATGATTT ATATTGCGCT GAATCATCTA GATAAAATTG	2340
25	TTTGTCTACA TGTTCAGGTG CTGTCTCAG TGCATCAAAC GCAGTTATAC TGCCAGTACT	2400
	TTGCTCATAA TATAGCAAAT ACCCGCCACC ACCAATACCT GATGCAAATG GTTCTACCAC	2460
30	ATTCAATGCC AGTTGAATTG CAATCACTGC ATCCATGGCG TTGCCACCTT GATCTAATAC	2520
	ATCCTTACCA ATTTTAGCCG CAAGAGGATG TGATACGGAA ATTAACCCTT CTTTAGATGT	2580
	TTTTGTCTGT TTGTCATTTA AGTTAATGAC CATACTATAT CCTCCTACTT TCTGT TAAAT	2640
35	ATTTAAAAA TTATTGATTA ATGGCTTTTT CTACTTTTTT TAAATCTTGA CGTTGCTCGT	2700
	TACCAGTATC GACAAGTGGT GTAATCGGTG ATGCAATTTT AAATTTATCG CCACGATAAA	2760
	ACTTAATAAA TTGATCCTGA TCTATCGCAT TAACTACTGC TTGTCTCAAG TTTGGATGCG	2820
40	TCTTAAATAT ACCTTTTTTA ATATTTAGCA TTAAAAAGAC TGACTTGCGT CCATTTTTGC	2880
	GAATAATGCT TAAATTTTTA TCCGACTTAA TTAAATCAAA ATGTTTTTGA TTCACATCTG	2940
	CCAACATATC AATTGAATGA TTTCTAAGTT CTGACAATGC ATTATTCGGG TCACCATTAA	3000
45	ACTTCAATGT AATATTTTTA ATTTTAGCTG GTCCATAACT ACCTTTTTCT GTTTCGTTGA	3060
	ATCCTGGATT ACGTTGAAAC GTTGCTTGAT ATGCATTTTT CTGTGTCATA ATGTATGCGC	3120
50	CACTTGATA CAGCGCATTT TTCCCATCTG AATTTGCAGG AATTGTACTG CTATCCCAT	3180
	ATCCTTTTGG ATATTCTTGA TTTACTTGAT TAACAAATTT TTTAGATAAA ATGCCTGCCG	3240
	AAGAGTGTGT TAAGTAATTT ACCTCTCGAG GCATCGATTG ATCTGTCGTA ATTTTAACAA	3300

55

	TATAAGCTTT AATCAACTTA TCATAGATTG ATTTATCGTC CTTGTCTTTC TCTTTACGCA	3420
	ACTGATCGAT GTCCTCATCT TTTAATATCT TGATGTCATT TATATGTTTG TGCATATTGT	3480
5	AAGTATTATT GTTAGGCACA GACTTTTTTAT CACGTGCTCT ATCTAAAGAA AACTTAACAT	3540
	CTTCAGCCGA TACACGCTCT CCAGTATTAC GTGCTTGTC ATTGACCACT TTCGCAAAAT	3600
	AATCATCATC TCTTAACAAG AAATAAAATG CTTTATTGTC CTTATTACACA GCATAATCAT	3660
10	GACTTAACGA ACCTTTCGTT GTTAAATGAT CATTTTCATC TAATAATAAT AACCTTGTGT	3720
	ACATATTATC ATTAATTGAA TATACTGACG GCGCAATTGA ACGTATTGGA TCCAATGTAG	3780
	GAATTTTACC ATCTTGTGTG GTCATCACAA GTGGCCGCGT ATCTCGTTCT CTACTATTGT	3840
15	TGTAATCAAA TTGTTGCCAT ATTAATGCAC GTGAATTTGG CAATCCAACA CTATTTTTAT	3900
	CTAACACTTT ATTGTCATAT ACTAAATTCT TTTTGTATCC ATATAAAGGC GCCATATACC	3960
20	CTTTATCAAA TACAACCTCA TCTTCAATTT GCTTATATGT TTGTTTAAACA TCTGCTTCAT	4020
	TTTGAGTAGA AGCTTTATTT AACAACTGGT CTACATGTTT ATCTTTCAAT AACTATTGTG	4080
	ATCCTGTAGA ACTAATAAT GCCGTCATAG CATAGTTCGG GTCACCAAAC ACTGTCATCC	4140
25	AGTCATCAAT TTGGATATCA TAATTGCCGG CTTGACGTTG TGTACGATAG CTACCATAAT	4200
	CTGGTTGGAT ATTCATCTTC ACGTTAAATC CTGCATTTTC CAATTGATCT TTAACGATAT	4260
	TCATATCATT TTCATAACTT GCTTGTCTTA GGAAATGTAT TGTGGTTCGC TCGCCTTTCA	4320
30	CTTCAACTTT CGATGACTTT TGAGCCACTT CTGATTTCTG AGGGACACCA CAACCACTTA	4380
	ATACCAACGC TAAAACTATA ATTGCGATAC TAATGATTTT CTTCACATCT ATCCCTACCT	4440
	TTTTAATGAA TTCTTGGATC TAGTGCATCA CGCACTGCAT CACCTATAAA ATTAAATGCT	4500
35	AAAACGACGA ACATAATACA AACACCAGGT ACAATAGCTA AATTACTGTG CGTTTCCAAG	4560
	TAGTTACTAC CGGTACGTAA AATGTTGCCC CATTCAGCTA CATCAGGTGC AACACCAAGT	4620
40	CCTAGGAAAC TTAAACTACT TGTGTGTTAAT ACAACCACAC CTATATTAA TGAAAAACGT	4680
	ACAATCATAG GCGCAATCGC ATTCGGTAAA ATATAACGCC ATATGATATT CCAAGTGTTT	4740
	TCACCAGTGA TACGTGCTGC ATCTACATAT TCCATGCGTT TAATTTCTAA AACACTGGCA	4800
45	CGCATTGTCC GTGCAAATGA TGGTATATTA CCGATACTTA AAGCAATAAT TAAATTTGGA	4860
	ATACTTGCTC CAAATGATGC AATAATTGCC ACCGCTAACA ATAATGATGG AATTGCAAAC	4920
	ACTACATCTA AAATTCGCAT TATTAAATTA TCAATATGAT TAAAAAACC TGCATAGTG	4980
50	CCTAGTAACA CACCAAAAAT AACTGCAATA ACTACTGAAA TAATTGAAAT TGAAAATGTC	5040
	AGCTTCGTTT CTACAACTAC GCGTGTAAT AAGTCTCTAC CGAAATCATC AGTACCAAAC	5100

55

	GTATCAAATG TAAATTGTGA CACAATTGAT AATGTCAGCA TGTAGACTAA AATAAGTAAC	5220
	CCGATAATCG CAATACGATG TCTAGTAGTT TTTTCGTATAA ACGATTCCCA CCCGTTATAA	5280
5	CTATGTATTT GCGATGTACG TTGGTAACGT CTAATACTTA CAAACATTAA TAATGTAAAT	5340
	ACGTTGCCTG TTAATGTCAT CAACAATAAC AACACTTCGA CGATACGTCG CCATAGGTCA	5400
	TGATGCTTCC ATGTTTGTTC CGTTGTTHAA ATAATAATTA AAATGATGGT TAAAACGATT	5460
10	AGCAATGTTT CAGCAATATA GAACGTATCG GCCACATAAC CTTTAAAAAG ATTTAATGCA	5520
	CTCGTTAATA TAACTAAAAT ATAAGTTGCT ATGGCGTAAC TTGCGAATAA TTTTAAGGAA	5580
15	GCTATCTTTG AATTAAGTTG TGCCATATGC CTCACTTCCT TTCGTTGATT TCACTACGTA	5640
	ATTTTGGATC GATTAAAGCA TAAAATATAT CAATAATTAA GTTTGCTAAA GATATTACAA	5700
	TTGATATATA TACGACCCCA CCCATGACTG CTGGAATATC AGGTATTAGT TGTTTTTGGA	5760
20	CGATATAACG CCCGATACCA TTAATGTTAA ATACTTGTTC CGTCACTGCT GAACCGCCTA	5820
	GTAACCTGTC CACTAGAAGA CCAACTAACG TTACAATTGG AATAATGGCA TTTTTCAAAA	5880
	TATGTTTAAT AACAACTTGT GTCGTCGATA ATCCTTTTGC ATAAGCAGTT AAAACATAAT	5940
25	CGctGCGCAT TACTTCAAGT ACAGAAGACC TTGTCATACG CGTGATAGAA GCAGCAATAC	6000
	TTGTTCCAAT GACAAGTACA GGTAAAATCA ACGATATTGG ATGTTCTGGC ATATAAGATG	6060
	GTGGCAAAAT ATCCAATTTT AATGAGAACG CTAAAATGAA TAATAGCCCT TGCCAGAAAC	6120
30	TTGGAATAGA TAAACCAATT AATGCAATTA TCATTAAAGT GATATCAAGC CAACTATTTT	6180
	GCTTCATCGC ACTGATAATA CCAATTGGTA TTGCAATAAT TAATGCCACC ATTAGCGCTA	6240
35	ATACTGCGAC AATTATTGTA ATTGGAATTC TTTGCGCAAC TGCTTTAGTC ACAACCTCAT	6300
	TCCCTTTGTA AGTCGTACCT AAGTCAAAGG TAAAAACACC CTTGATGGTA TCCCACAATT	6360
	GAATZAAATA AGGTTCTGTA AGATGATGTA ATACATTGAA TTGATGTATC TGTGCCTTTG	6420
40	TTGCATTTTG TCCAGTATG CTATAAGCCG CATCAAGCGG TGAAAAATAC AGAATGGTAA	6480
	ACACACTGAC AATAACACCA ATGATGACAA TCACAGCCAT GACAATTCGT TCAAAAATAT	6540
	ATCTAACTAA TGGCTGTAAA TAAAAAGTCA ATAAGATGAA CATCGGCAAG GCCAATATCA	6600
45	CTTTGATCAT GATGAACTTA TGAAATAATA CATTTTCAAA GTATGTTGAA AAATGTGCTT	6660
	GTTCAATATT CTTTGAATC GTATTAGAAC TTTGTGCCTT GAATATTTTT AATGCTTCTT	6720
	TATGTATTTG TGTGGATGAC TTTTGCTGCG ATAAATATTT ATATTTTTGA TGTAACGCCT	6780
50	GTTCAATTTT TGAAATTTCA GAATTATTAG CGTAAAAATT TTTCTCTTA GCAGAAAAGA	6840
	AAAACTTTAT CACTGCATAT AAAAATATTG GCAAGCTTAA TACCGATAAT ACAAACCTGT	6900

55

	CTTGTAATAAT AATCTTGAGT AGATTACTAT GATATACAAA AGTATAGAAT AAATTTACAC	7020
	ATTTGTGaAT AGGGAGGCAC AACATCATGT CAAATTTATT AGAAGTCAAC AGTCTGAATG	7080
5	TACAATTCAA TTATGATGAA ACTACAGTTC AAGCGGTAAA AAACGTCTCT TTCGAATTAC	7140
	GAAAAAACA TATCCTAGGT ATTGTTGGTG AATCAGGATC AGGAAAAAGT ATTACCGCTA	7200
	AATCTATTTT AGGGCTACTA CCAGATTATC CAGATCACAC ATTAACAGGA GAAATTATTT	7260
10	TTAATGGGCA ATCGTTAAAT AATTTATCAA CTTCAGCGTT ACAACAAATT CGAGGTAAGG	7320
	ATATTTCAAT GATTTTTTCAA GATCCACTCT CTTCGTTGAA TCCAAGATTA ACGATTGGCA	7380
15	AACAAATTAC AGAAGTAATA TTTCAACATA AACGTGTATC TAAATCTGAA GCAAAGTCGA	7440
	TGACAATAGA CATTTTAGAA AAAGTAGGTA TAAAACATGC AACTCGACAA TTTGATGCTT	7500
	ATCCACATGA ACTTTCTGGT GGTATGCGTC AACGTGTCAT GATAGCAATG GCATTGATTT	7560
20	TAAAGCCACA AATTTTAATC GCAGATGAaC CAACAACGGC ATTAGATGCC AGTACACAAA	7620
	ATCAATTACT GCAGTTAATG AAGTCCCTTT ATGAGTACAC AGAAACATCT ATTATTTTTA	7680
	TCACTCACGA TTTAGGCGCT GTGTATCAAT TTTGCGACGA TGTGATTGTA ATGAAAGATG	7740
25	GAAGTGTCTG TGAAAGTGGC ACGGTTGAAA GTATTTTTAA ATCGCCACAA CATACCTATA	7800
	CAAAACGCTT AATAGATGCG ATTCCTGATA TTCATCAAAC GCGTCCGCCA AGACCGTTAA	7860
	ACAATGATAT TTTATTAAAA TTCGATCGCG TGAGyGgGAT TACACATCAC CGAGTGGCAG	7920
30	CCTATACCGA GCAGTTAATG ATATTAACCT GGCTATTAGA AAAGGCGAAA CATTAGGCAT	7980
	TGTCGGTGAA TCAGGGTCAG GGAAATCGAC ATTAGCTAAG ACGGTCGTCG GTCTAAAGGA	8040
35	AGTGTGAGAA GGCTTTATTT GGTATAACGA ATTACCATTA AGTTTATTTA AAGATGATGA	8100
	ATTGAAATCT TTACGACAAG AGATACAAAT GATTTTTCAA GATCCATTCTG CATCTATTAA	8160
	TCCAAGATTT AAAGTCATTG ATGTGATTAA ACGACCACTA ATCATTTCATG GGAAAGTCAA	8220
40	AGATAATGAT GACATTATTA AAAGTGTCTG ATCGTTGTTA GAAAAGGTTG GCCTAGATCA	8280
	AACTTTCTTA TATCGCTATC CACACGAATT ATCTGGTGGG CAACGTCAGC GTGTAAGTAT	8340
	CGCGAGAGCA CTTGCTGTTG AACCTAAAGT GATTGTTTGC GACGAGGCAG TGTCCGCTTT	8400
45	AGACGTTTCA ATTCAAAAAG ATATCATCGA GTTATTAAAA CAATTACAGT TAGACTTCGG	8460
	CATCACTTAT TTATTCATCA CACATGACAT GGGTGTATC AATGAAATAT GTGATCGCGT	8520
	TGCAGTTATG AAAAAATGGCG AAATCGTTGA ACTGAATAAC ACAGAAGATA TTATCAAACA	8580
50	TCCGCAGTCA GACTATGCAA AGCAACTTAT TTCAGAAGTA GCAGTTATTG CTAAATAAAA	8640
	GTCAATGCGTT GTGCAACTTT ATCACTGTAT GGTCTGAAAT AAATTGCGCG ACTTCTGATG	8700

55

5 TATCAAGTTT TAGGTGCTTT GCCATGATTT AAGAGTCACC CCCATACTTT GGGCATTTTA 8820
 ACGCCAGAAT AAATCCCCCG CCACTATGTG AAGTGTGGGG GATTATTTAT ATTTTATTAG 8880
 AATATTGAGA TTTTGTAGTG TGTCAACTTA GCTTAGTCAA TGTATATTTA ACGTCACTTA 8940
 CTCTTTTCTT TTCATAATTA ACACATTCAA ATAACTTTG ATCAAAAAAC ACAAAGTTAA 9000
 10 AAGTACCATC TTGTAATATG CTCTCATACA TTATCCCGTC ATATTTAAGG CTTCGAATAT 9060
 AATCAGCTAA ATATTGAAAT GGCAAATAAT CTATTCCTTG TTCATCGCTT GGATTGTGTA 9120
 TTCCTTTATG AATCTTTTTT AATGTTGGT AATTTACAAA ATACTTTCTA AATCCATCAT 9180
 15 CGCCAGCTTT GATTGCATTA CTAGTTAAAT TAGTTAAATT CGCAATTTTC AATTTCTCTT 9240
 TTGTCACGTT TTTTGTAAAC TTAACCTTAC CTATATAAAT AATGTCATTA TGCTTAGGTT 9300
 TAACTTCTTC TATACTGACC TGTCTTTTTG TACTAAGGTA TAATACGCTT ATCCATTTAG 9360
 20 AATTCAATCT TCCTGCCGTT GCAAATCCCT TTGGTGGTGA CATTAGTTCA CTTTTCTCTG 9420
 TAATGAACTT AACTATTCTA GATCTATATA ATGGTTCAAA TCTTTCTCTA AATTCCTCAA 9480
 TACTATAGTA ATTAGTAGTG ATATCGAGAA AGAACGCTAA ATTCTCTAAA TTGATCATAT 9540
 25 TTTTATGAAA TCTATTTTAA TACTTCAAGC TCTCACAAAA TCCATCCCAG TCATTATTTG 9600
 CTACAATTAG ATTTTATTTT GTATATTTTT TATCGTTTAT GATTTTAGCG CCTACTAAAT 9660
 CTTCCAACAC TCGTCTATCT AAATTTTCAT CATCTTTAAA AAGTTCATT AAAATACAAC 9720
 30 TTATTTGAGC TTCCTCAACA TTAAATATAC TCCAGTCGTC TTTTAATGCT ATTTCAATCT 9780
 TTTTACCTTC TTTTGGGCTA AAAGTATCTG GTAAATTTAT ACTAATATCA TATAATTCTA 9840
 35 ATGCTGGTCT TAAATAATCT CTAATAAGTT CTAATTTATC TATGTCCTTA GTCGTATCAA 9900
 ATATTTTAAC ACCAAGATGA TTGTTATCAA TATCACAATT GTCAAATTTG CTATTTATCA 9960
 TTTGCAATGA TTTCTACGAT TTCAGTATTA TTAAACATT TTTACATAT TTTCATTTTG 10020
 40 AGACTCCAAG TATCTATTCA TAATTTCTAG GTGATGCATG ATAGATAACC TTTTAATTAA 10080
 ACCTAATCCT GGATaCTTAT TATTTTCATT TAATTCCTCA AATTGTCCCA AGCGCATAAG 10140
 45 ATCTATTTTT AATATCTAAG TTTTTTGACC ATGTTACTAA TT 10182

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3491 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

	AACTCAGGCA ATTGAAACAG CATTAGGTGC TTCATTACAA CATGTCATTG TAGATTGAGA	60
	AAAAGATGGA CGCCAGGCTA TTCAATTTTT AAAAGAACGT AATTTAGGTC GTGCGACGTT	120
5	TTTACCATTA AATGTTATAC AGAGTAGAGT GGTAGCGACT GATATTAAAT CTATTGCTAA	180
	AGAGGCAAAC GGATTTATTA GTATCGCTTC GGAAGCAGTT AAAGTAGCAC CAGAATATCA	240
10	AAATATTATC GGAATTTTAT TAGGTAATAC GATTATCGTT GATCATTTAA AGCATGCAAA	300
	TGAATTGGCA CGTGCGATTA AATATCGAAC TCGTATTGTT ACTTTGGAAG GTGATATTGT	360
	AAATCCTGGT GGtTCTATGA CTGGTGGTGG CGCTCGTAAG TCAAAAAGTA TTCTGTCTCA	420
15	AAAAGACGAG TTGACAACAA TGAGACACCA ATTAGAAGAT TACTTGCGTC AAACAGAATC	480
	ATTTGAACAA CAATTTAAAG AGTTGAAGAT AAAAAGTGAT CAATTAAGTG AACTGTATTT	540
	TGAAAAAAGT CAAAAGCATA ATACACTTAA AGAGCAAGTG CATCATTTTG AAATGGAGCT	600
20	CGATAGATTA ACTACACAAG AAACACAAAT AAAAATGAT CATGaAGAAT TCGAATTTGA	660
	AAAAAATGAT GGTtATACGA GTGACAAAAG TCGACAACT TTGAGTGAAA AAGAAACTTA	720
	TCTAGAAAGT ATTAAAGCAT CTTTAAAACG ACTAGAAGAT GAAATTGAAC GCTACACAAA	780
25	ACTTTCTAAA GAAGGTAAGG AAAGCGTTAC TAAAACACAA CAAACCTTAC ATCAGAAACA	840
	ATCTGATCTT GCTGTGGTTA AAGAGCGTAT TAAAACACAA CAACAGACAA TAGATCGATT	900
30	AAATAATCAA AATCAACAAA CTAAACATCA ATTAAAAGAT GTTAAAGAAA AAATTGCATT	960
	CTTTAATTCTG GATGAAGTGA TGGGCGAACA AGCTTTTCAA AATATTAAAG ATCAAATTAA	1020
	TGGTCAACAA GAAACGAGAA CACGCTTATC AGATGAATTA GATAAATTGA AACAACAACG	1080
35	TATTGAGTTG AATGAACAAA TCGATGCGCA AGAAGCTAAA CTACAAGTTT GTCACCAAGA	1140
	TATTTTAGCT ATCGAAAATC ACTACCAAGA TATTAAAGCT GAACAATCAA AGCTAGATGT	1200
	ATTAAATTCAT CATGCGATAG ATCATTaAAT GATGrATATC AATTGACTGT TGAACGTGCG	1260
40	ArATCTGAAT ATACGaGTGA TGraTCGATg ACGCATTACG TAAAAAAGTT AAGTTAATGr	1320
	AGaTGyCGAT TGATGrACTA GGTCTGTAA ACTTAAATGC AATTGAACAA TTTGAAGAGT	1380
	TAAATGAACG TTATACATTT TTAAGTGAAC AACGTACAGA TCTTCGTAAA GCTAAAGAAA	1440
45	CATTAGAGCA AATTATAAGT GAAATGGATC AAGAGGTTAC TGAAAGATTT AAAGAACTT	1500
	TCCATGCTAT TCAAGGACAT TTTACAGCTG TGTTCAAACA ATTGTTTGGT GGAGGCGATG	1560
50	CAGAATTGCA ATTAAC TGAA GCCGATTATT TAACAGCTGG TATTGATATT GTGGtACAAC	1620
	CACCGGGTAA AAAGTTGCAA CATTTATCGT TACTGAGTGG TGGTGAGCGT GCATTAACTG	1680
	CTATTGCTTT ACTATTTGCA ATTTTAAAAG TAAGATCTGC ACCTTTTGTt ATATTAGrTG	1740

55

	TATCAGACGA AACACAATTC ATTGTTATTA CACACCGTAA AGGAACAATG GAATTTGCAG	1860
	ATAGGTTATA CGGTGTAACA ATGCAAGAAT CAGGTGTTAC TAAACTTGTG AGTGTGAATT	1920
5	TAAATACAAT AGATGATGTG TTGAAGGAGG AGCAATAATG AGCTTTTTTA AACGCTTAAA	1980
	AGATAAGTTT GCAACAAATA AAGAAAATGA AGAAGTTAAA TCCTTAACAG AAGAACAAGG	2040
10	TCAAGACAAA TTAGAAGATA CACATTCTGA AGGTTCAACG CAGGACGCAA ATGATTTAGC	2100
	AGAAAATGCT GAAGTGAAAA AGAAGCCACG CAAGTTGAGT GAAGCGGATT TTGATGACGA	2160
	TGGCTTAATA TCAATTGAAG ATTTTGAAGA AATTGAAGCT CAAAAAATGG GTGCTAAATT	2220
15	TAAAGCAGGA CTCGAAAAAT CTCGTCAAAA TTTCCAAGAA CAATTAAATA ATTTGATAGC	2280
	GAGATATCGT AAAGTAGATG AAGACTTTTT TGAAGCTTTA GAAGAAATGT TAATCACTGC	2340
	AGACGTCGGT TTAAATACAG TGATGACGTT AACTGAAGAA TTACGTATGG AAGCACAACG	2400
20	ACGTAATATT CAAGATACTG AAGATTTGCG TGAAGTCATT GTTGAAAAGA TCGTAGAGAT	2460
	TTACCATCAA GAAGATkATA ATTCAGAAGC TATGAACTTA GAAGATGGTC GTTTAAATGT	2520
	CATTTTAATG GTTGGTGTGA ATGGTGTGG TAAAACAACA ACAATTGGAA AATTAGCTTA	2580
25	CCGATATAAA ATGGAAGGTA AAAAAGTAAT GTTAGCTGCG GCGGATACTT TTAGAGCGGG	2640
	TGCTATTGAT CAATTGAAAG TTTGGGGCGA ACGTGTGGT GTAGACGTAA TTAGCCAAAG	2700
	TGAAGGTTCT GATCCAGCTG CTGTTATGTA TGATGCgATT AATGCCGCTA AAAACAAAGG	2760
30	TGTTGATATT TTAATCTGTG ATACCGCTGG ACGTTTACAA AATAAmACAA ATCTAATGCm	2820
	AGAATTAGAA AAAGTTAAGC GTGTAATTAA TCGAGCAGTG CCAGATGCGC CTCATGAAGC	2880
35	ATTACTATGT TTAGATGCTA CAACTGGTCA GAATGCGTTG TCACAAGCTA GAAACTTTAA	2940
	AGAAGTAACA AATGTTACAG GTATTGTATT AACGAAATTA GATGGTACAG CCAAAGGTGG	3000
	TATCGTATTA GCCATTCGTA ATGAATTGCA CATCCCAGTT AAATATGTAG GTTTAGGTGA	3060
40	GCAATTAGAT GACTTACAAC CATTTAACCC TGAAAGTTAT GTCTACGGCT TATTCGCTGA	3120
	TATGATTGAA CAAAATGAAG AAATAACAAC AGTTGAAAAT GATCAAATTG TAACAGAAGA	3180
	AAAGGACGAT AATCATGGGT CAAAATGATT TAGTtAAAAC GTTACGAATG AATTATTTGT	3240
45	TTGATTTTtAT CAATCCTTAT TGACGAATAA ACAACGTaAT TATTTGGAAT TATTTTATCT	3300
	TGAAGATTAT TCTTTAAGTG AAATCGCAGa TACTTTTAAT GTGAGTAGaC AAGCAGTTTA	3360
50	TGATAATATA AGAAGAACTG GCGATTTAGT TGAAGATTAT GAAAAGAAAT TGAATTATA	3420
	CCAGAAATTT GAGCAACGCC GAGAAATATA TGATGAAATG AAACCACATT TAAGTAATCC	3480
	AGAACAAATA C	3491

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4253 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

10 AGTACGTTTT ATAATTATAA GTACGTAATT AACATATTAA CATATCGCAA GSTATGTATTT 60
 AAATAAGATT GTTATAATTT CAAAGTTCAT CCAAGATTAT GGCGTTTGCA TTTACCTATT 120
 15 AAAAACGTTA TTATATCAAA GATGCGAAAG ATAATACGGG TTTATTTTAT GAAAGTGAGA 180
 AGGATAAAAT GGATAATGAG CAACGCTTAA AAAGAAGAGA GAATATAAGG AATTTCTCGA 240
 TTATAGCACA TATTGACCAC GGAAAATCTA CATTGGCTGA TAGAATTTTA GAAAATACCA 300
 20 AATCAGTTGA AACAAGAGAT ATGCAAGATC AGTTACTAGA TTCAATGGAT TTAGAAAGAG 360
 AACGTGGTAT TACAATCAAA TTAAACGCgT ACGTTTAAAG TACGAAGCTA AAGATGGAAA 420
 TACTTATACA TTCCATTAA TCGATACGCC TGGACACGTC GATTTTACAT ATGAAGTGTC 480
 25 ACGTTcTTTG GCAGCTTG TG AGGGCGCGAT TTTAGTAGTA GATGCGGCTC AAGGTATCGA 540
 AGCACAAACA TTAGCAAATG TTTATTTAGC ATTAGATAAT GAGTTAGAGT TATTGCCTGT 600
 TATTAACAAA ATTGATTAC CTGCTGCAGA ACCTGAACGC GTGAAACAAG AAATTGAAGA 660
 30 TATGATAGGT TTAGACCAAG ACGATGTTGT TTTAGCAAGT GCTAAATCTA ACATTGGAAT 720
 TGAAGAGATA CTAGAGAAAA TAGTTGAAGT TGTGCCAGCT CCAGATGGTG ACCCAGAAGC 780
 35 ACCACTAAAA GCGTTAATAT TTGATTCTGA GTATGATCCA TATAGAGGGG TAATTTTCATC 840
 GATAAGAATT GTGGACGGTG TTGTTAAAGC CGGAGATAAA ATTCGAATGA TGGCCACTGG 900
 TAAAGAGTTC GAAGTAACAG AAGTTGGAAT TAATACACCT AAGCAGCTTC CAGTTGATGA 960
 40 ATTAACAGTT GGTGATGTTG GTTATATTAT TGCAAGTATT AAAAATGTTG ATGATTCTAG 1020
 GGTGGTGAC ACCATCACAT TAGCTAGTAG ACCTGCATCA GAACCATTGC AAGGTTATAA 1080
 GAAAAATGAAT CCAATGGTAT ATTGCGGACT GTTCCCAATA GATAACAAAA ATTATAATGA 1140
 45 TTTAAGAGAA GCATTAGAAA AATTACAATT GAATGATGCA TCATTAGAAT TTGAGCCTGA 1200
 ATCGTCACAA GCATTAGGTT TTGGTTATAG AACTGGTTTC TTAGGTATGT TACACATGGA 1260
 AATAATTCAA GAAAGAATTG AAAGAGAATT TGGTATTGAA TTAATTGCAA CTGCACCATC 1320
 50 TGTAATTTAT CAATGTGTTT TAAGGGACGG TTCAGAAGTG ACGGTTGATA ACCCAGCACA 1380
 AATGCCAGAT CGTGATAAAA TTGATAAAAT ATTTGAGCCA TATGTTCTGT CAaCTATGAT 1440

	TATAAATATG GACTATTTAG ATGATATTCG TGTAAATATT GTTTATGAAT TACCTTTAGC	1560
	TGAAGTTGTA TTTGATTTCT TCGATCAACT TAAATCTAAT ACTAAAGGAT ATGCATCATT	1620
5	TGATTATGAA TTCATCGAAA ATAAAGAAAG TAATTTAGTC AAGATGGATA TTTTATTAAA	1680
	TGGTGATAAA GTGGATGCGC TAAGCTTCAT AGTTCATAGA GATTTTGCAT ATGAACGTGG	1740
10	TAAAGCATTG GTTGAAAAAC TTAAAACGTT AATTCCAAGA CAGCAATTG AAGTACCTGT	1800
	ACAGGCTGCA ATAGGACAAA AAATTGTAGC GCGTACAAAT ATTAAATCAA TGGGTAAAAA	1860
	CGTTTTAGCT AAATGTTATG GCGGTGACAT AAGCCGTAAA CGTAAATTAC TTGAAAAACA	1920
15	AAAAGCAGGT AAAGCTAAGA TGAAAGCAGT TGGTAATGTT GAAATCCAC AAGATGCTTT	1980
	CTTGCTGTA TTGAAAATGG ATGATGAATA ATTTTAAAA ATCAATTAAC AATTTACAAT	2040
	GAATAAAGTT TAATAACTAA AAAGAGGGAG CCTAGGATAA ATTAACGTCC TGGGCTTTAC	2100
20	AATGTTATAT TGGCAGCCAT CGACAGAGTT AAAATGAGCT TATAACAATG GGGCCCCAAC	2160
	ACAGAAGCTG ACGAAAAGTC AGCTTACTAT AATGTGCAAG TTGGGGTGGG GCCCCAACAT	2220
	AGAGAATTTT GAAAAGAAAT TCTACAGGCA ATGCAAGTTG GGGTGGGACG ACGAAATAAA	2280
25	TTTTGCGAAA ATATCATTTC TGTCCCACTC CCTTATGCAT GAGTTTTACT CATGTAATTT	2340
	TATTTTAAAG GACATATTAC ATCTGGCTAA TGTGTAAGAG CCACTACATA ATAAATCATT	2400
30	AGTGGTTCTT TATTATTTCT ATCTCACTCC CTCTAAACAA GAATAAATAT TAAAATGAAT	2460
	CGATATATTA GACAATCATT GATTAAACGT TAAAGTAAA AGTAAGAATA ATTGCAGATA	2520
	GTCCAACAGG ATATAGCCGA TTGGATAAAA AGTCTGAGAA GCGGGGCATT AAAATGACGG	2580
35	TACAAAGTGC ATATATACAT ATTCCATTTT GTGTAAGAAT ATGTACATAT TGTGATTCA	2640
	ATAAATATTT TATACAGAAT CAACCTGTAG ATGAGTACTT AGATGCACTA ATCACAGAAA	2700
	TGTCTACAGC AAAATATAGG ATCTTAAAGA CCATGTATGT AGGTGGCGGC ACACCAACGG	2760
40	CCCTTTCTAT TAATCaGTTG GAAAGATTAC TTAAAGCAAT ACGTGATACG TTTACAATCA	2820
	CAGGCGAGTA TACATTTGAA GCAAATCCTG ATGAGTTAAC TAAAGAGAAA GTCCAACAT	2880
	TAGAGAAATA TGGAGTAAAA AGGATTTCAA TGGGCGTTCA AACATTCAAG CCGGAGTTAT	2940
45	TGTCTGTTTT AGGTAGAACG CACAATACTG AAGATATTTA CACTTCGGTG TTAAATGCTA	3000
	AAAACGCAGG TATTAAATCA ATCAGTTTAG ATTTAATGTA TCATTTACCG AAACAGACGA	3060
50	TTGAAGATTT TGAACAAAGT TTAGATCTAG CTTTAGATAT GGATATTCAA CATATTTCGA	3120
	GTTACGGCTT AATACTTGAA CCTAAAACCC AATTTTATAA TATGTATAGA AAAGGCTTGC	3180
	TCAAACCTGC TAATGAGGAT TTAGGTGCTG ACATGTATCA GTTGCTGATG TCTAAGATAG	3240

55

AACATAATAA GGTTTACTGG TTTAATGAGG AATATTATGG ATTTGGAGCA GGTGCAAGTG 3360
 GTTATGTAGA TGGTGTGCGT TATACGAATA TCAATCCAGT GAATCATTAT ATCAAAGCTA 3420
 5 TAAATAAGA AAGTAAAGCA ATTTTAGTAT CAAATAAACC TTCTTTGACT GAGAGAATGG 3480
 AAGAAGAAAT GTTTCTTGGG TTGCGTTTAA ATGAAGGTGT GAGTAGTAGT AGGTTCAAAA 3540
 10 AGAAGTTTGA CCAATCTATT GAAAGTGTCT TTGGTCAAAC AATAAATAAT TTAAAAGAGA 3600
 AGGAATTAAT TGTAGAAAAG AACGATGTGA TTGCACTTAC AAATAGAGGG AAAGTCATAG 3660
 GTAATGAGGT TTTTGAAGCT TTCCTAATAA ATGATTAAAA AAAATTGAAA TTTCGAGTCT 3720
 15 TTAACATTGA CTTACTTTGA CCAATTTGAT AAATTATAAT TAGCACTTGA GATAAGTGAG 3780
 TGCTAATGAG GTGAAAACAT GATTACAGAT AGGCAATTGA GTATATTAAA CGCAATTGTT 3840
 GAGGATTATG TTGATTTTGG ACAACCCGTT GGTTCATAAA CACTAATTGA GCGACATAAC 3900
 20 TTGAATGTTA GTCCTGCTAC AATTAGAAAT GAGATGAAAC AGCTTGAAGA TTTAAACTAT 3960
 ATCGAGAAGA CACATAGTTC TTCAGGGCGT TCGCCATCAC AATTAGGTTT TAGGTATTAT 4020
 GTCAATCGTT TACTTGAACA AACATCTCAT CAAAAACAA ATAAATTAAG ACGATTAAAT 4080
 25 CAATTGTTAG TTGAGAATCA ATATGATGTA TCATCAGCAT TGACATATTT TGCAGATGAA 4140
 TTATCAAATA TATCTCAATA TACAACCTTA GTTGTTTCATC CTAATCATAA ACAAGATATT 4200
 ATCAATAATG TACACTTGAT TCGTGCTAAT CCTAATTTAG TTATAATGGT TAT 4253
 30

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3395 base pairs
 35 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TCCCTAATCG AACAAAATTA TGCGCATAAA CAAAGTAGAT TGATATAAAA TTCTTAATTA 60
 TCAGAAATATA TTTACAAATC TGAATTTTAT TAGTATATTG GrTAGTrTTC ATAGAGGCAT 120
 45 GACGGTaTTT GAGCAGGATT TTAAATCGGg ATTTTATAAT CGATTTAAGA GAGGCCActT 180
 TGCTTGcACA TTAATACTGT cAATGGGAGG GGAATGTATA TGAGTrAAGC ACATCAATTA 240
 ATTCAAGAGG ATGAACATTA TTTTGCGAAA TCAGGACGTA TTAAATATTA TCCGTTAGTG 300
 50 ATTGATCATG GATATGGAGC AACATTGGTT GATATTGAGG GGAAGACATA TATCGATTTG 360
 TTATCGAGTG CGAGTTCTCA AAACGTAGGT CATGCACCTA GAGAAGTAAC AGAAGCGATA 420

55

	GTACGTTTAG	CTAAGAAGCT	TTGTGAGATT	GCACCTGGAG	ATTTTGAAAA	AAGAGTGACC	540
	TTCGGATTAA	CCGGATCAGA	CGCAAATGAT	GGCATCATT	AATTTGCCAG	AGCATATACA	600
5	GGGCGTCCTT	ATATCATTAG	TTTCACTAAT	GCATATCATG	GTTCAACTTT	TGGCTCATTG	660
	TCTATGTCAG	CTATTAGTTT	AAATATGCGC	AAACATTATG	GTCCGTTATT	GAATGGTTTT	720
10	TATCATATTC	CGTTTCCAGA	TAAATATCGT	GGTATGTACG	AGCAGCCACA	AGCTAATTCA	780
	GTAGAAGAAT	ATTTAGCACC	CTTAAAAGAA	ATGTTTGCGA	AGTATGTACC	TGCTGACGAA	840
	GTAGCATGTA	TTGTTATTGA	AACGATACAA	GGCGATGGTG	GACTTTTAGA	ACCAGTTCCA	900
15	GGGTATTTTG	AAGCGTTAGA	AAAGATTTGT	CGTGAACATG	GTATTTTAAAT	CGCTGTCGAT	960
	GATATTC AAC	AAGGTTTTGG	GAGAACAGGT	ACATGGAGTT	CAGTCTCGCA	TTTTAATTTT	1020
	ACGCCTGATT	TAATCACTTT	CGGAAAATCC	TTAGCAGGTG	GTATGCCTAT	GTCAGCAATT	1080
20	GTTGGACGCA	AAGAGATTAT	GAATTGTTTA	GAAGCACCAG	CACATTTATT	TACAACAGGT	1140
	GCTAATCCAG	TTAGTTGTGA	AGCTGCATTA	GCCACAATTC	AAATGATTGA	AGATCAGTCG	1200
	CTTCTTCAGG	CTAGTGCGGA	AAAAGGGGAA	TATGTTAGGA	AACGAATGGA	TCAATGGGTA	1260
25	TCTAAATACA	ATAGTGTAGG	CGATGTTAGA	GGTAAAGGTC	TGAGCATTGG	TATTGATATT	1320
	GTTTCCGACA	AAAACTCAA	AACACGTGAT	GCCAGTGCGG	CACTTAAAT	TTGTAATTAC	1380
30	TGCTTTGAGC	ATGGCGTAGT	TATTATAGCT	GTAGCAGGAA	ATGTGTTGCG	ATTCCAACCG	1440
	CCATTGGTAA	TAACATATGA	GCAATTAGAC	ACGGCGTTAA	AACTATAGA	AGATGCACTG	1500
	ACTGCTTTGG	AAGCAGGTAA	CTTAGATCAA	TATGACATAT	CTGGACAAGG	TTGGTAATAG	1560
35	CGATTATCTT	AATATAAAAT	AAAAAATCAT	TTCCACATCT	GGATGTTAAT	CAGATGGGAA	1620
	ATGATTTTTT	TTATTTTTTA	TTTTGGTGGG	TGGTATTCAG	CTACGTCATT	TTTCTTAGAA	1680
	TGTGTAAGTC	CATAACTTAA	ATATAGGATG	ATACCAACAA	TAAACCAAAT	TAAAGTGTAT	1740
40	AATTCGCTT	CGAATCCTAA	TCCCAGAAAT	ACTAGCAATA	CTAAAACAAA	TGTAATTGCT	1800
	GGTAACACAG	GATATAAAGG	TAATTTAAAT	GCAGGAATTG	GTAGATCTTT	ACCTTcACGC	1860
	TTTCTCAAAC	GATACATTGC	TAATGAAACG	AACATAAATG	CAACAAGTGT	ACCTGCTGAA	1920
45	ATTAATTGTG	CTAAAAATGC	GAATGGGAAC	ATAGAACCAA	TTAAAACACC	AATAATAGTA	1980
	AGTATAACTA	GTGCGCGATT	AGGTAAATGT	TTGTCGTTTA	AGTGGCTTAA	CCATGAAGGT	2040
50	AATAAGCCGT	CACGTCCAAA	TGAATAAAGT	AAACGTGAGC	CTGCTAACAT	CATACCAATT	2100
	AATGCTGTAA	ACATACCGAT	AACAGAGATA	GCTTGAACAA	TAGCTGCTAC	AACACCATGA	2160
	CCACTTTGAC	GTAAAGCCCA	ACCAACAGGT	TCAGCATTGT	TTGCGTATTG	TGAGTAATGG	2220
55							

CCAAGAATAC CTCTAGGCAT TGTCTTTTGA GGATCAAGTG CTTCTGCTGA GTTTGCTGCG 2340
 ATAGAATCGA AACCGATATA CGCTAAGAAA ATCATTGAAA CACCAGCATA TATGCCTTGC 2400
 5 CATCCACCAA AGTCACCTGT AGCAGTTACT TTGTGTTCTG GAATAAATGG CACATAGTTA 2460
 CTAACATTTA TTGCTGTAA ACCTACGATG ACAAATAAAA TAATAGCTAA TACTTTTAAA 2520
 ATAATAAAA TATTTTCCAT ACGAGCTGCT TCCGACATAC CACGTGATAG TAATAATGCA 2580
 10 GTTAATAAAA TAACGATAGC AGCAATAATA TCGATAAAAC CGCCATTGT ACCAAATGGA 2640
 TTTGATAATG CTGCAGGTAA TTCGATGCCA ATTGGTTTCA CAAGTCCGCG TAAATTCGCT 2700
 15 GAGAATCCTG ATGCAACAAA GGCTACGGCG ATAAATATT CAGCTAATAG AGCCCAACCG 2760
 GCAACCCATC CAAAAATTC ACCAAATAAT ACATTGACCC AAGAATAGGC TGAACCTGCA 2820
 AATGGCATAG CGGCAGCCAT TTCTGCATAA GTAAATGCAA CTAAACCAGC AACAATAGCA 2880
 20 GCGAGTAAGA ATGATAACGC AACGGCCGGT CCTGCATGTT CTGCAGCAAC AATGCCAGGT 2940
 AGCGTAAAGA TAGATGTCGA TACAATTGTT CCTACACCTA AAGCTAAGAA ATCACGCACC 3000
 CGAAGTGATC GCTTTAAATG ACCATCTTTA TTTTGATAGA TAGCCGGATC CTCTTTTCGT 3060
 25 GCTATTTTAT TGAAAAACT TCCATAAAC TTTCTCCCA AACATTCATA AACAATTCTA 3120
 TACGGTGTTT TTTAATATGT TATATCATAG CACAAATAAT CAATATTTTG TCTAAAAATT 3180
 CTGAAAAATC ACAACTTTAT GTTACGTATT AATGACTTGT CTTGATAACA TCCATAGATT 3240
 30 TTTTAAATGA TAAAACTGAT TATAACAGAT ATTAAATGAA TAAGTACTAT TTTTGCnAA 3300
 TTTTCTAACA ATTTTGCACA TTATATGTTT AAAATCAATT TCATGTTTAT GGTCTGATTG 3360
 35 GCTAGTGTGT ATGAAATGTA AntCTTTGAC TnnGA 3395

(2) INFORMATION FOR SEQ ID NO: 120:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13508 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

ATCAGGTAAT GCCATGCGTT TAGCTGAAAA TTTTTCAGA ACGTTTAAGT GATATCGGAC 60
 ATCAAGTTGT TTTGATGTCA ATGGATGAAT ATGATACGAC AAACATCGCG CAGTTAGAAG 120
 50 ATTTATTTAT TATTACGTCT ACTCATGGTG AAGGAGAACC GCCTGATAAT GCATGGGATT 180
 TCTTTGAATT TTTAGAAGAC GATAACGCAC CTAATTTAAA TCATGTGAGA TATTCAGTAC 240

TACTAGAAAA TCTAGGCGCT GAGCGTATAT GTAAGCGTGT AGATTGTGAT ATTGATTATG 360
 AAGAAGACGC AGAAAAGTGG ATGGCAGACA TCATTAATAT TATTGATACC ACATCAGAAG 420
 5 GTATTCAAAG TGAATCGGTG ATAAGTGAAT CAATTAAGTC TGCCAAAGAA AAGAAATATT 480
 CTAAATCAAA TCCATACCAA GCAGAAGTAT TAGCGAATAT CAATTTAAAT GGTACCGATT 540
 CAAATAAAGA AACACGACAT ATAGAATTTT TACTTGATGA TTTTAGTGAA TCATATGAAC 600
 10 CAGGAGATTG TATAGTAGCA TTACCGCAAA ACGACCCTGA ATTGGTTGAA AAATAATAT 660
 CCATGTTAGG TTGGGATCCG CAATCTCCGG TGCCAATTAA TGATCATGGT GATACAGTTC 720
 CTATTGTTGA AGCACTAACA TCACATTTTG AATTTACTAA ATTAACATTG CCATTATTGA 780
 15 AAAATGCAGA TATCTATTTT GACAATGAAG AATTATCTGA ACGTATTCAA GATGAGTCAT 840
 GGGCGCGTGA ATATGTTATA AATCGGGACT TTATAGATTT AATAACAGAT TTTCCAACTA 900
 20 TAGAATTACA ACCTGAGAAT ATGTATCAAA TCCTTAGAAA ATTACCACCA AGAGAGTATT 960
 CGATTTCTAG TAGTTTTATG GCAACGCCAG ATGAAGTGCA TATTACCGTT GGTACGGTTC 1020
 GTTATCAAGC ACATGGACGT GAGAGAAAAG GTGTATGCTC GGTTCATTTT GCTGAGCGAA 1080
 25 TTAAACCAGG CGATATAGTA CCAATTTATT TGAAGAAAAA TCCGAACCTC AAATTTCCGA 1140
 TGAAGCAAGA TATACCGGTT ATTATGATTG GACCAGGTAC TGAAATTGCT CCTTTTAGAG 1200
 CATATTTACA AGAACGTGAA GAACTTGGA TGACTGGAAA AACATGGTTG TTCTTTGGTG 1260
 30 ATCAACACCG TAGTTCTGAC TTTTATATG AAGAAGAAAT AGAAGAATGG CTTGAAAATG 1320
 GAAACTTAAC ACGCGTAGAT TTAGCATTTT CAAGAGACCA AGAACACAAA GAATATGTAC 1380
 AGCATCGTAT AATGGAAGAA AGTAAACGTT TCAATGAATG GATTGAGCAA GGCGCACAAT 1440
 35 CTATATTTGT GGCGATGAAA AATGTATGGC GAAAGATGTC CATCAAGCCA TTAAAGATGT 1500
 ATTGGTAAAA GAACGTCATA TTTCTCAAGA AGAAGCAGAG TTATTATTGC GACAAATGAA 1560
 40 ACAACAACAA CGCTATCAAC GTGATGTTTA TTAGCGATTG GTGTAAATA TTTTAAGGTG 1620
 TAATGATGTA AAAAGATATA AAGGATGTTG CTCAACATGA ATATGCCATT AATGATAGAT 1680
 TTAACAAATA AAAATGTCGT CATAGTTGGT GGAGGCGTCG TTGCAAGTCG TCGGGCACAA 1740
 45 ACATTAAATC AATACGTTGA ACATATGACG GTCATCAGTC CGACAATCAC TGAAAACTT 1800
 CAAAATATGG TAGATAACGG TGTCGTCATA TGGAAGAAA AAGAATTTGA ACCAAGCGAT 1860
 ATTGTAGACG CGTATCTAGT TATTGCAGCA ACCAATGAGC CACGTGTCAA TGAAGCGGTA 1920
 50 AAAAAAGCCT TACCTGAGCA TGCCCTTTTT AATAATGTTG GAGATGCATC AAATGGCAAT 1980
 GTTGTATTTT CAAGTGCACT ACACCGCGAC AAGCTAACTA TCAGTGTATC AACTGATGGT 2040

55

	TACAGTTCGT ATATCGACTT TTTATATACT TGCCGACAGA AAATAAAAGT ACTTGATATA	2160
	ACATATAACG AAAAGCAACA GTTACTGTCA CAAATTGTGT CACAAGAATA TTTAAATCAT	2220
5	GACAAACAAG CTCAATTTTT AGCGTGGTTG GATGTAAGAT AATAATAGCG GACCGTCTAA	2280
	CCGTCTAAGG TAAGTCTTCT TATTTTAACT TTAACGCTTA ATCATTGAAA TTAAGACATG	2340
10	GGCGGCTTTG TGAATAGTCT AATAATGAAG GATTTAAGCG ATAATGATAT GCGTTTTAAA	2400
	TATGAATATT ACAATAGAGA AAAAGATACG TAGAACAAAC TTAATAAAAT AGGTGGATAA	2460
	ATTGAAATCT GGTGAAGTC GTTACTATCA TAGCGACCTT TAGCCAGATT TTTTGTGCAA	2520
15	TAGAAAGCAA TAATAAAAT GATAGATCAA AATGAAATAC AGGACAGGAT ATACAAGGAT	2580
	TAGTCATGCC ATGTTATCAA GTAGGAAAAT CAACTTCAC TATTGATAGT TACGCAAAAA	2640
	AGATTTTTTT GATAAAATGA GATAACTTAA ATATAAAAA TTATATTAAT TATAATATTT	2700
20	AAGTTAAAGA GGGGATTAT GTAAATTGTA TTTAAAGTGG AGGGAGAAAA TAATATGAAT	2760
	AGTGATAATA TGTGGTTAAC AGTAATGGGG CTCATTATTA TTATTTCAAT TGTAGGTTTA	2820
	CTCATTGCCA AAAAGATAAA TCCAGTTGTA GGTATGACAA TCATACCTTG CTTAGGGGCA	2880
25	ATGATTTTAG GATATAGTGT GACAGATTTG GTTGGATTTT TTGCTAAAGG GTTAGATCAA	2940
	GTCATCAACG TTGTTATTAT GTTTATCTTT GCCATTATTT TCTTTGGCAT CATGAACGAT	3000
	AGTGGTTTAT TCAAGCCGCT TGTCAAACGC TTAATATTAA TGACACGAGG CAATGTCGTC	3060
30	ATTGTCTGTG CAATGACAGC TTTAATTGGC ACAATAGCCC AATTAGATGG GGCCGGTGCG	3120
	GTAACATTTT TGCTTTCTAT TCCTGCATTA TTACCTTTAT ATAAAGCGTT AAATATGAAT	3180
35	AAATATTTAT TGATTTTACT ATTAGCATT AAGCGGGCGA TTATGAACAT GGTACCTTGG	3240
	GGAGGTCCAA TGGCTCGTGT AGCTGCAGTG TTAAAAGCCA AAAGTGTCAA TGAATTATGG	3300
	TATGGATTAA TACCTATTCA AATAATAGGT TTCATTCTTG TTATGTTGTT TGCGGTATAT	3360
40	CTTGGATTTA AAGAACAGAA ACGTATCAAA AAAGCAATAG AGAGAAATGA ATTACCGCAA	3420
	ACACAAGATA TAGATGTACA TAAATTAGTT GAAGTATATG AACGAGATCA AGATGTAAGG	3480
	TTTCCTGTAA AAGGACGTGC AAGAACAAAA TCATGGATAA AATGGGTGAA TACAGCTTTA	3540
45	ACTTTAGCTG TTATTCTATC GATGTTAATA AATATTGCGC CACCTGAATT TGCATTTCATG	3600
	ATAGGTGTTy CGTTGGCACT TGTTATTAAT TTTAAATCAG TGGATGAACA AATGGAACGA	3660
	TTAAGAGCgC ATGCGCCGAA TGCATTAAATG ATGGCTGCAG TGATTATTGC AGCAGGTATG	3720
50	TTTTTAGGTG TACTAAATGA AACCAGGTATG CTTAAAGCGA TTGCGACCAA TTTAATCAAA	3780
	GTGATTCCTG CAGAAGTAGG ACCATACTTG CATATTATTG TAGGTTTACT TGGCGTACCA	3840

55

	ACAGCAGGGC AATTTGGTGT ACCGTCTGTA TCAACAGCTT ATTCAATGGT CATAGGGAAT	3960
	ATTATAGGTA CATTTGTCTAG CCCATTTTCA CCAGCCTTAT GGTTGGCAAT TGGTTTAGCA	4020
5	GAGGCAAACA TGGGCACGTA TATTAAGTAT GCATTCTTTT GGATTGGGG ATTCTGCTATC	4080
	GTTATGTTAG TAATTGCAAT GTTGATGGGC ATTGTGACGA TTTAAGTATG AAAAAATAGA	4140
	AACTATGGTC ACGTTGCAAA ATGAAATAAT AGTTGCATAA ACATGTCGAA ATGACGGACG	4200
10	AATCTTTAAA CAATTTTAAA AATTAATGAA ATAATTGTGT AGAAATATGA ATTTCACTAA	4260
	ATGTTAATAA CTTTGTGACG TTTTAGTTAA CAGACTAATA AAAATTGAA AATACTATAT	4320
	ATAGTGGTAT AACGTAATGA GTAGACACAA TATATAGGAA GAAGGGGTAA AATGAATCAA	4380
15	ATCGAAGAAG CATTAAACGGG TTTGATTTCT AAAGATCCTG CTATTGTTAA CGAAAATGCT	4440
	AACAAAGATA GTGATACATT TTCAACAATG AGAGATTTAA CAGCAGGTAT CGTTTCTAAA	4500
20	TCTTACGCAT TAAATCATTT ATTACCAAAG CACGTTGCAG ATGCACATCA AAGAGGGGAC	4560
	ATACATTTTC ACGACTTAGA TTATCATCCA TTCCAACCGT TAACTAACTG TTGTTTAATA	4620
	GATGCTAAAA ATATGCTACA TAATGGATTT GAAATAGGCA ACGCGAATGT AACTTCACCA	4680
25	AAATCAATAC AAATGTCATC AGCGCAGCTT GTACAAATTA TAGCCAATGT TTCTAGCAGT	4740
	CAATATGGTG GCTGTACGGT TGACCGCGTT GACGAATTAC tTAGTACATA TGCACGACcA	4800
	TAATGAAGAA CAACATAGGA ATATsCGCAA AGCAATTTGT CAAAGAATCT GAAATTGATC	4860
30	GTTATGTTGA TCAACAAGTC ACTAAAGACA TCAATGATGC GATTGAAAGT TTAGAATATG	4920
	AAATTAATAC CTTATATACA TCTAATGGAC AGACACCTTT TGTAACATTA GGATTTCGGCT	4980
	TAGGTACAGA TCATTTAAGT CGCAAAATTC AACAAGCTAT CTTAAATACT CGTATCAAAG	5040
35	GCTTAGGAAA AGACCGCACG ACAGCGATTT TCCCGAACT TGTATTTTCA ATTAAAAAAG	5100
	GAACCAACTT TAGTCCGCAA GATCCGAACT ATGACATTAA ACAACTAGCA TTAAAGTGTT	5160
40	CAACGAAACG TATGTATCCA GATATTTTAA ATTATGACAA ACTCGTAGAA ATATTAGGTG	5220
	ATTTCAAAGC GCCAATGGGT TGTCGTTTCAT TTTTACCAAG TTGGAAAGAT GCGGAAGGTC	5280
	ATTTTGAAAA TAATGGTCGT TGTAATCTTG GTGTTGTTAC ACTTAATTTA CCTAGAATGG	5340
45	CATTAGAATC TGCCGGTAAT ATGACGAAAT TCTGGGAAAT CTTTTATGAA CGTATCGATG	5400
	TGTTACATGA TGCATTACTT TATCGTATAA ATCGTTTGAA AGATGCTGTA CCGAATAACG	5460
	CACCGATTTT ATATAAAAGT GGCGCATTTA ACTATAAATT AAAAGAAACA GATGATGTTG	5520
50	CTGAGTTATT TAAAAATAAA CGTGCAACGA TTTCAATGGG CTATATAGGG TTGTATGAAA	5580
	CAGCTACTGT TTTCTATGGT CCAGACTGGG AAACATCTCA AGAAGCAAAA GCATTTACGC	5640

55

	GGTTCAGTAT TTmCAGTACG CCGAGTGAAT CGCTAcGGAT CGTTTTTGTC GTTTAGACCA	5760
	AGAGAGATTT GGAGATATTA AAGACATTAC AGATAAAGGA TATTATCAAA ACTCTTTCCA	5820
5	TTATGATGTA CGTAAAGATG TTACACCTTT TGAAAAGTTA GATTTTGAAA AAGATTATCC	5880
	TTATTATGCG AGTGGTGGTT TCATTCACTA TTGTGAGTAT CCGAAATTGC AACACAATTT	5940
10	GAAAGCACTA GAAGCGGTAT GGGACTACTC TTATGACAAA GTTGGTTACT TAGGTACAAA	6000
	TATTCGATT GATCATTGTT ATGAATGTGA TTACGATGGA GATTTTGAAG CAACTGAAAA	6060
	AGGATTTAAA TGCCCGAACT GTGGCAATGA TAATCCTAAA ACAGTTGATG TCGTTAAACG	6120
15	AACATGTGGT TACCTAGGCA ATCCAGTTCA ACGTCCAGTA ATTAAAGGCC GTCATAAAGA	6180
	AATTTGCGCA CGAGTAAAC ATATGAAAGC GCCTAAAGAA TGATACTTTT AGACATTAAA	6240
	CAAGGACAAG GTTATATTGC TAAAATAGAA TCAAATAGCT TTGTTGACGG TGAAGGAGTA	6300
20	AGATGCAGTG TTTATGTATC AGGATGTCCA TTTAATTGTG TTGGATGTTA TAACAAAGCC	6360
	TCACAAAAGT TCAGATATGG CGAGAAATAC ACTGATGAAA TATTAGCAGA AATATTAGAT	6420
	GATTGCGATC ATGATTATAT ATCTGGGCTA AGTCTATTAG GTGGCGAACC ATTTTGTAAAT	6480
25	TTGGATATTA CATTAAATCT TGTCAAAGCA TTTGAGCAC GTTTTGGAAT TACAAAGACA	6540
	ATTTGGGTAT GGAATGGATT TTTATATGAA TATTTAGCAA ATGATTGTAC AGAACGTCGA	6600
30	GAGTTATTAT CATAATTGA CGTTTTAGTA GATGGTCTAT TTATACAACA CTTATTCAAA	6660
	CCTGATTTAC CATATAAAGG TTCTTTAAAT CAACGCATTA TAGATGTACA ACAATCACTC	6720
	TCGCATGCCG GTATGATTGA ATATATAGTT AGTTGAATAT GTATTAGAAG TCAAGGTAAC	6780
35	ATTCGTTGCC TTGGCTTCTT TTTAGGTTAG GTACATAATT GAAAGTTAAT AAAAGCAATT	6840
	CTTTATAAAA ATATATTGAT AGAATATGAC CTAACAATCA TTTTGATACC AATACTAAAA	6900
	GTTGATATC CGTTTTTTAA AAAAGTTGAA AGAGAAAAGT GGTATTTTAG TGGGAAGGAA	6960
40	GTCIAACTTT TTGGTAGCGT TTTACAATAA ATAAATATTC GTTAATAACG TATAAATATT	7020
	CTTAAATGCC ATTCTAGTAA AATTTGTTAA ATTCGTTAAA TCGTAACTTA AACTGTATT	7080
	TTTAGCGCTA TTAAGGTTTT GTTTATTACG GGAAAAATTA TATAAATATT CAATAATTGC	7140
45	CAAGTTTCAA ATTGTATGAA ATTTGCATTA TTATTAAATG TTAGTTATTG TCAATTTTGT	7200
	GAATCAATAT AATTATTACA TTTTGAGATA AATCGAAACA GGATTCATAA AATTAATAAT	7260
	TAGGGGGAGC ACAATTGAAA AAAGAGAAAG TTATGGACTG GACGACCTTT ATAGGGACAG	7320
50	TAGCTGTACT TCTTTTTGCA GTTATACCTA TGATGGCTTT TCCAAAAGCA AGTGAAGATA	7380
	TCATCACTGG TATTAATAGT GCCATTTCTG ATTCAATTGG TTCGATATAT TTATTTATGG	7440

55

	TTGGTAAAGC AAGTGATAAA CCAGAATTTA ATACATTTAC ATGGGCGGCA ATGCTGTTTT	7560
	GTGCAGGCAT AGGCTCTGAT ATTTTATACT GGGGCGTTAT TGAATGGGCT TTTTACTATC	7620
5	AAGTTCCACC AAATGGGCGG AAAAGTATGA GTGATGAAGC ACTCCAATAT GCGACGCAAT	7680
	ATGGTATGTT CCACTGGGGG CCAATTGCTT GGGCTATTTA TGTTCACCA GCATTACCAA	7740
10	TTGGTTATTT AGTATTTGTT AAAAAACAAC CGGTGTATAA AATTAGTCAA GCTTGTGCTC	7800
	CGATTTTAAA AGGTCAAACA GATAAATTTG TAGGTAAAGT TGATAGATATC TTATTTATCT	7860
	TTGGATTGCT AGGTGGTGCG GCAACATCAC TAGCGTTAGG TGTGCCATTA ATTTCTGCAG	7920
15	GCATAGAAAG ATTAAGTGGT TTAGATGGTA AAAATATGAT TTTACGTTTCG GCCATTTTAT	7980
	TAACAATCAC GGTATATTTT GCCATTAGTT CATATACAGG ATTGAAAAAA GGTATTCAAA	8040
	AGTTAAGTGA TATCAACGTT TGGCTATCCT TTGTACTTTT AGCCTTTATA TTTATTATTG	8100
20	GACCGACTGT TTTTATTATG GAAACGACAG TGACAGGGTT CGGAAATATG TTGAGAGATT	8160
	TCTTTCATAT GGCAACATGG TTAGAACCAT TCGGTGGTAT TAAAGGTCGA AAAGAAACGA	8220
	ATTTCCACCA AGACTGGACA ATATTCTACT GGTCATGGTG GTTAGTATAT GCGCCATTTA	8280
25	TCGGTTTATT TATCGCTAGA ATTTCAAAG GTCGACGCTT TAAAGAAGTC GTGCTAGGAA	8340
	CAATTATTTA TGAACGCTT GGATGCGTAT TATTCTTTGG TATTTTTGGT AACTATGCTG	8400
30	TGTATTTACA AATTTCTGGA CAGTTTAATG TAACACAATA TTAAATACA CATGGTACAG	8460
	AGGCAACCAT TATTGAAGTG GTGCATCATT TACCATTCCC ATCATTGATG ATTGTACTAT	8520
	TCTTAGTATC TGCTTTCTTA TTCTTAGCAA CAACATTTGA TTCGGGTTCA TATATTTTAG	8580
35	CGGCAGCATC TCAGAAAAAA GTGGTAGGCG AACCATTACG TGCCAATCGT TTATTCTGGG	8640
	CATTTGCATT GTGCTTATTG CCATTTTCAT TGATGCTAGT TGGTGGTGAA CGTGCATTAG	8700
	AAGTATTGAA AACTGCTTCA ATACTGGCAA GTGTGCCATT AATTGTTATT TTTATTTTCA	8760
40	TGATGATATC ATTTTAAATC ATTTTAGGGC GCGATAGAAT TAAACTTGAA ACGCGTGCTG	8820
	AAAAATTAAA AGAAGTTGAA CGTCGTTTAT TGCGAATCGT TCAAGTATCa GAAGAAGAAC	8880
	AAGACGATAA TTTATAATTC AAAGCGGGTC TGGGACGACG AAATGaATT TGTGAAAATA	8940
45	TCATTTCTGT TCCaTTCCCC TTTTTTTAGT AGCATTGTAG GATGAACTTT TAGGTTTTCA	9000
	TTAATGTTGT ACTAAAAGAT TTAATTTTTT AGTGCTCCAA GTACTTATTT ATTGTATGAA	9060
50	GCATATTCTA AATCGAAGTT TGAAAGACTC TCATTGATTA TTAAATTAAA TAAAGGGTAT	9120
	GCGTATGTAC AATTCAAATT AATCGAAGGA TGAAATAAAA TGACTAATCA ATTTAAAAAT	9180
	AAACAGTCCA AATTACATGA CAGTTTAGAA TCCATCACAA AAAACTTATA TGCGACACCT	9240

55

	ACAGAATATT GTTATCTATC ATTCCGGACA CTTAGGTGAC TCCCAACAAG ACATTGCATC	9360
	ATTAGGTGGT GTTTCAAAAG TATTGATGAA TCATGATCAT GAATCTATAG GAGGTTCTAA	9420
5	TCAAGTTGAA GCCCCTTACT TTATACATGA AAATGATGTG GCTGCACTGA AACATAAGAT	9480
	TTCTGTTCAA AAACAATTTA GTAATCGTGT AATGTTGGAT AAGGATTTAG AAGTTATTCC	9540
10	CGCGCCTGGA CATACACCAG GGACGACACT ATTTTTATGG GATGATGGTC ATCACCCTTA	9600
	CTTATTTACT GGAGATTTTA TATGTTTTGA AGGGAAGAGA TGGCGTACAG TTATATTAGG	9660
	TTCAAGTGAT AGAGAAAAAT CTATTCAAAG TTTAGAGATG GTTAAAGAAT TAGATTTTGA	9720
15	TGTACTTGTA CCTTGGGTTA CTATCAAAGA TGAACCGTTA GTTTATTTTG TAGAAAATGA	9780
	ATATGAAAAA CGTGAACAAA TACAAAATAT TATTGATAGA GTACGTGAGG GCGAGAATAG	9840
	CTAATTGAAA TATATTGGCG AAgCAATGTA ACGAATCTAA GAAAGCCCTA GAAAATACCT	9900
20	CCATAATTGA TTGTCATATA AAACAAAAAC GGTAATTTCT ATTTATTGAG ATAGAAATTA	9960
	CCGTTTATTT CGTGGACCTA TTGCATTGTT TTTATCATGC ATAATCATCA TTGTCGTTGT	10020
	TTGAGTCAAT TTAAATTTTC AGAATCAGAA GGCTGTTCTG GAATTGGGAA ATATTTGAAA	10080
25	ATTTACCCGC TTTCAATCGC TTCGGTTAAC TGTTCTAACC ATTCGTAATA AACATGTGTA	10140
	TGATCAAGCT GAGCTTTAAT TTTTGTGCC TCTTGTGTTT CAGCTTCAGT TAAATCACTG	10200
30	CTTTCAAGTA ATGGATTGAT AATAGCTTGA GCATCTTTTA CTGCTTCGAC ATTGATGTCA	10260
	ATTTACGCT GGAATTTTTT AGTGAAAAAG TTTCGGAAAA AGATGAAAAA GTCTTTCTCG	10320
	GCGATAAAAT GTTGTTTGCG GCTTCCTCTC GTAAATTGTT GTTTAACAAT ATCAAATTCC	10380
35	TGCAATTTCT TAACGCCAGC ACTCATACTT GGTGTTGCTCA TTTGCAATTG ATGACGCATT	10440
	TCATCAAGCG TCATACTGCC TTCAAACACC ATTGTGCCAT ATAAGTTTCC TACACTTCTA	10500
	TTAGTGCCAT ACAAATCCAT TGTCTGTCCA ATTGAATTAA TTACAATATC TTTTGCTTGT	10560
40	TCTAATTGTT GCTGTTTGTT CTGAGAACGA GTCATCATTG CACCTCCGTA CATCATTTTG	10620
	GTCACGTTAA AATAAATACT AATACATTAT AAAACCTTTT CTAAAAAAG ACATTAAAAA	10680
	TATTTAAAGC ATTAAAGTTA AATGTTTCGT TAAATAAAAA TCTAACGAAC TTACAAAAC	10740
45	TAATTCTTGA GTTGTTTTGT AAATTGACAC ATTTTTCATT TCTATGCTAA CATAAGTnTG	10800
	TAAAATTcGT TAAATAAAAA TTTAACAAAC TTAACGGrGG TTGTTGAAkG GrACTTTTAA	10860
50	aACATTTATC TCAGCGTCAA TATATTGATG GTGAGTGGGT TGAAAGCGCG AATAAAAAATA	10920
	CAAGAGATAT TATCAATCCT TACAATCAAG AAGTGATATT TACGGTTTCT GAAGGGACAA	10980
	AAGAGGATGC AGAACGTGCA ATCTTAGCTG CAAGACGTGC GTTTGAGTCT GGTGAATGGT	11040

55

	AACATCgCGA AgCgTTAGCA CGATTAGAAA CATTAGATAC TGGAAAAACG TTAGAAGAAT	11160
	CATATGCAGA TATGGATGAT ATTCATAATG TGTTTATGTA TTTTGCTGGA TTAGCAGATA	11220
5	AAGACGGTGG CGAAATGATT GATTCACCAA TTCCAGATAC AGAAAGCAA ATTGTTAAAG	11280
	AACCAGTAGG TGTAGTTACA CAAATTACAC CTGGAATTA TCCGTTATTA CAAGCATCAT	11340
10	GGAAAATTGC GCCAGCGCTT GCTACGGGTT GTTCACTAGT TATGAAACCA AGTGAAATTA	11400
	CACCATTAAc AACAAATACGT GTTTTTGAAT TAATGGAAGA AGTTGGTTTC CCTAAAGGAA	11460
	CAATTAATCT TATTCTAGGT GCAGGTTCTG AAGTTGGTGA CGTAATGTCA GGTCATAAAG	11520
15	AGGTTGACCT TGTATCATTT ACAGGTGGCA TTGAGACTGG TAAGCATATT ATGAAAAATG	11580
	CTGCTAATAA TGTTACGAAT ATTGCCTTGG AACTTGGCGG TAAAAATCCA AACATTATCT	11640
	TTGATGATGC TGATTTTGAA TTGGCAGTAG ACCAAGCGTT AAATGGTGGA TATTTCCATG	11700
20	CAGGTCAAGT TTGTTcAGCA GGATCAAGAA TATTAGTACA AAACAGTATT AAAGACAAAT	11760
	TTGAGCAAGC ACTTATTGAT CGCGTGAAAA AAATCAAATT AGGTAATGGT TTTGATGCTG	11820
	ATACTGAAAT GGGACCAGTG ATTTCAACAG AACATCGTAA TAAGATCGAA TCTTATATGG	11880
25	ATGTAGcTAA AGCAGAAGGC GCAACAATTG CTGTTGGTGG TAAACGTCCA GATAGAGATG	11940
	ATTTAAAAGA TGGTCTATTC TTCGAGCCAA CAGTCATTAC AAATTGTGAT ACGTCAATGC	12000
	GTATTGTACA AGAAGAGGTT TTCGGACCTG TCGTTACTGT AGAAGGCTTT GAAACTGAAC	12060
30	AAGAAGCGAT TCAATTAGCG AATGATTCTA TATATGGTTT AGCAGGTGCT GTATTTTCTA	12120
	AAGATATTGG AAAAGCACAA CGCGTTGCTA ACAAGTTGAA ACTTGGAACG GTGTGGATTA	12180
35	ATGATTTCCA TCCATATTTT GCACAAGCGC CATGGGGTGG ATACAAACAA TCAGGTATCG	12240
	GTAGAGAATT AGGCAAAGAA GGCTTAGAAG AGTACCTTGT TTCAAACAC ATTTTAACAA	12300
	ATACAAATCC ACAATTAGTG AATTGGTTTA GCAAATAAAA ATTAGATAAG GTGAGTGCCA	12360
40	TTGTAAGAAC ACAAGACACT CACTTTGTTT TGTATAAGTG GCGAAATGTT GATTGATAAT	12420
	TTGGAATAAA CGCAAAATGA ATCATAGATT ATTTCAATTAC TGTTAGTAAC AATCGTAAAA	12480
	GGAAAAGCGA GTGTTTTGGT TAGCTAAGTT TAGCAATTCA ACGATAACCA ATCAGCCACT	12540
45	AACAAATATT TCATGCAATA CTCACTTTGA AATACAACAA ACTTTGGAGG TCATAACGAT	12600
	GAGTAACAAA AACAAATCAT ATGATTATGT CATCATTGGA GGAGGCAGTG CAGGTTCTGT	12660
	ACTAGGTAAT CGTCTGAGTG AAGATAAAGA TAAAGAACTC TTAGTATTAG AAGCGGGTCG	12720
50	CAGTGATTAT TTTTGGGATT TATTTATCCA AATGCCTGCT GCGTTAATGT TCCCTTCAGG	12780
	CAATAAATTT TACGATTGGA TTTATTCAAC AGATGAAGAA CCACATATGG GCGGTCGTAA	12840

55

TCAACGTGGT AATCCAATGG ACTATGAAGG CTGGGCAGAA CCAGAAGGTA TGGAAACTTG 12960
 GGATTTTTCG CACTGTTTAC CGTATTTTAA AAAATTAGAA AAAACATACG GTGCAGCGCC 13020
 5 TTATGATAAA TTTAGAGGCC ATGATGGACC AATTAAGTTA AAACGAGGGC CAGCAACGAA 13080
 TCCTTTATTC CAGTCATTCT TTGATGCAGG TGTGAAGCA GGCTATCATA AAACACCTGA 13140
 TGTGAATGGA TTTAGACAAG AAGGTTTTGG ACCGTTTCGAT AGTCAAGTAC ATCGTGGTCG 13200
 10 CCGAATGTCA GCTTCAAGAG CATATTTTACA TCCAGCGATG AAGCGTAAAA ACTTAACCGT 13260
 TGAAACACGT GCCTTTGTAA CTGAAATTCA TTATGAAGGT AGAAGAGCAA CTGGTGTTAC 13320
 GTATAAGAAA AATGGCAAAC TACATACCAT CGATGCTAAT GAAGTCATTT TGTCTGGTGG 13380
 15 GGCATTCAAT ACGCCACAAT TACTACAATT ATCTGGTATC GGTGATTCAG AGTTCCTAAA 13440
 ATCAAAAGGC ATTGAGCCAC GTGTTTATTT ACCTGGTGTG GGTGAAAAC TGAAGATCA 13500
 20 CTTAGAGG 13508

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7646 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTAAGTATTG TCTTGATTTC CTAATAAAGT TATATCTTGT AATTCATCTT GTTGACGGCC 60
 ATGTGCCATA TAAAGCGCTC CTTTAAATTT ATTTTTTTAT TATTTTGGCG TCTCGGCGTG 120
 35 CTTTTTCAAA CATGTAATAA CTTGCACCGA TAATAACGAC GTAACCTAAT GTTGCATAGA 180
 AATCTGGAGA TTCTCCGAAT AGAATAAATC CAACTATTGC TGTGAAAATT ATAGATGCAT 240
 40 ACGTAAAAAT AGAAATATCT TTTGCTGCTG CAAAACATA TGCTAAAGTA ACACCAATTT 300
 GACCCACAGC GGCAGCTAAG CCAGCCCCTA ATAGATAAAG TATTTGCATC TGACTCATTG 360
 GTTCATAAGT ATATGCAGTG AAAGGTATTA AAACGATGAC AGAAAATAAG GAGAAGTAAA 420
 45 ATACTATAGT ATATGGTGCT TTTCTGTGAC TAAGTGCTCG AACACATGTA TATGCTGATG 480
 CTGCAAAAAT ACCTGAGAAT AAGCCAGCTA ATGATGGAAT CATAGATGAT GAAAATTCAG 540
 GTTCACTAT TAAGAGCAaC CTAAAATAGC AATTATCATT GCTGTAATTT GaTACTTCCT 600
 50 TACCTTTTCA TGtAAGAAA CAATGCTTaa TAAAATCGTC CAGAAAGGAT TGAGTTTCAT 660
 TAATGAATCG GCATCACTAA GTACCATATG ATCAATGGCA TAAATATTTA ACAATACACC 720

55

	TGGCTGATGG TATTTATATA TAAAAAATAA TGGAATAAAC ATTGCTACTA AGTTTCGTGC	840
5	TAATGATTTT TGAAAAACAG GAAGGTCACC TGCAAGTCTG AAAAACAACG ACATAAAACT	900
	GAAACCAATA GCCGAAATTA AAATGGCAAT GATACCTTTT ACTTTAGGAT TCAATTTTAT	960
	CGCCTCTTTT ATATAAAATT AACGTATTTA TATTAGCATA AAACAACATG TTGTGCATAA	1020
10	ATAGTTGAAA TTTACTATAA AAAGACTATA ATAGACTGTA GCGAACAAAC GTTCTGTGTT	1080
	TATTTGTCGG AATAATAGGG CATTACACTT TTATGAATGT TTGTGTTATT ACATAAAACA	1140
	AATATCAATT CAGTATCAAG CTAATAAGCT TTTTCTTGAT TTCTGTTGAT ACAATTGAGA	1200
15	TTGACACAGA TTTAAAAAAA TCAAGTGATA TCTACTAAAA AATTTTTTTA AATTTGTTCA	1260
	AGTTTTTCTA ATTTAGTATT GGTGCCTAGT TGGAACGTTT TACGAACATT CGATTAGAAA	1320
	ATGGCACTTT AAATCATAGT GTGTCTTATG TATAATGAAA CACATAATAT AGTGTGGTG	1380
20	AAACGAAAAA gACACAATAT CTTGTGTTTT GTATGCAAAAT GCTTTATTTA TGAAGAAATT	1440
	ACATTTAAAA GTAATTTAAC ACAGAAATTT AATAGTTATT ATCAATTAAT AGTCATATTT	1500
	TTAGAAAATG TACTGAGCAA ATGGAAGATA TCCAATGATG TAAACACTAC ATATAGTGAT	1560
25	TTTTATACAT TCAACCCATA TAAGCTACTA TTTTCTCAA TATAAATCTA TGCAATTGGT	1620
	TTACATTTGA GAAAATAAGT AGCTTCATTA TAGTTAATAC AATGCTGAGA TAACCATAGT	1680
30	AACCATGTTG TTAAAGCATT TTTTAATTGG AATGACTACT TTATTTAAAA GGGTTGAAGA	1740
	AAGAAGGTGA TCCAATGAAA ATAATATATT TTTCATTTAC TGGAAATGTC CGTCGTTTTA	1800
	TTAAGAGAAC AGAACTTGAA AATACGCTTG AGATTACAGC AGAAAATTGT ATGGAACCAG	1860
35	TTCATGAACC GTTTATTATC GTTACTGGCA CTATTGGATT TGGAGAAGTA CCAGAACCCG	1920
	TTCAATCTTT TTTAGAAGTT AATCATCAAT ACATCAGAGG TGTGGCAGCT AGCGGTAATC	1980
	GAAATTGGGG ACTAAATTTT CAAAAGCGG GTCGCACGAT ATCAGAAGAG TATAATGTCC	2040
40	CTTTATTAAT GAAGTTTGAG TTACATGGAA AAAACAAAGA CGTTATTGAA TTTAAGAACA	2100
	AGGTGGGTAA TTTTAATGAA AACCATGGAA GAGAAAAAGT ACAATCATAT TGAATTAAAT	2160
	AATGAGGTCA CTAAACGAaG AGAAGATGGA TTCTTTAGTT TAGAAAAAGA CCAAGAAGCT	2220
45	TTAGTAGCTT ATTTAGAAGA AGTAAAAGAC AAAACAATCT TCTTCGACAC TGAAATCGAG	2280
	CGTTTACGTT ATTTAGTAGA CAACGATTTT TATTTCAATG TGTTTGATAT TTATAGTGAA	2340
50	GCGGATCTAA TTGAAATCAC TGATTATGCA AAATCAATCC CGTTTAATTT TGCAAGTTAT	2400
	ATGTCAGCTA GTAAATTTTT CAAAGATTAC GCTTTGAAAA CAAATGATAA AAGTCAATAC	2460
	TTAGAAGACT ATAATCAACA CGTTGCCATT GTTGCTTTAT ACCTAGCAAA TGGTAATAAA	2520

55

	ACATTTTTTAA ACGCAGGCCG TGC GCGTCGT GGTGAGCTAG TGTCATGTTT CTTATTAGAA	2640
	GTGGATGACA GCTTAAATTC AATTAACTTT ATTGATTCAA CTGCAAAACA ATTAAGTAAA	2700
5	ATTGGGGGCG GCGTTGCAAT TAACTTATCT AAATTGCGTG CACGTGGTGA AGCAATTAAA	2760
	GGAATTAAAG GCGTAgCGAA AGGCGTTTTA CCTATTGCTA AGTCACTTGA AGGTGGCTTT	2820
10	AGCTATGCAG ATCAACTTGG TCAACGCCCT GGTGCTGGTG CTGTGTACTT AAATATCTTC	2880
	CATTATGATG TAGAAGAATT TTTAGATACT AAAAAAGTAA ATGCCGATGA AGATTTACGT	2940
	TTATCTACAA TATCAACTGG TTTAATTGTT CCATCTAAAT TCTTCGATTT AGCTAAAGAA	3000
15	GGTAAGGACT TTTATATGTT TGCACCTCAT ACAGTTAAAG AAGAATATGG TGTGACATTA	3060
	GACGATATCG ATTTAGAAAA ATATTATGAT GACATGGTTG CAAACCCAAA TGTGAGAAA	3120
	AAGAAAAAGA ATGCGCGTGA AATGTTGAAT TTAATTGCGC AAACACAATT ACAATCAGGT	3180
20	TATCCATATT TAATGTTTAA AGATAATGCT AACAGAGTGC ATCCGAATTC AAACATTGGA	3240
	CAAATTAAAA TGAGTAACTT ATGTACGGAA ATTTTCCAAC TACAAGAAAC TTCAATTATT	3300
	AATGACTATG GTATTGAAGA CGAAATTAAA CGTGATATTT CTTGTAACTT GGGCTCATT	3360
25	AATATTGTTA ATGTAATGGA AAGCGGAAAA TTCAGAGATT CAGTTCCTC TGGTATGGAC	3420
	GCATTAAGT TGTGAGTGA TGTAGCAAAT ATTCAAATG CACCAGGAGT TAGAAAAGCT	3480
	AACAGTGAAT TACATTCACT TGGTCTTGGT GTGATGAATT TACACGGTTA CCTAGCAAAA	3540
30	AATAAAATTG GTTATGAGTC AGAAGAAGCA AAAGATTTTG CAAATATCTT CTTTATGATG	3600
	ATGAATTTCT ACTCAATCGA ACGTTCAATG GAAATCGCTA AAGAGCGTGG TATCAAATAT	3660
35	CAAGACTTTG AAAAGTCTGA TTATGCTAAT GGCAAATATT TCGAGTTCTA TACAACCTCAA	3720
	GAATTTGAAC CTCAATTCGA AAAAGTACGT GAATTATTCG ATGGTATGGC TATTCCTACT	3780
	TCTGAGGATT GGAAGAACT ACAACAAGAT GTTGAACAAT ATGGTTTATA TCATGCATAT	3840
40	AGATTAGCAA TTGCTCCAAC ACAAAGTATT TCTTATGTTT AAAATGCAAC AAGTTCTGTA	3900
	ATGCCAATCG TTGACCAAAT TGAACGTCGT ACTTATGGTA ATGCGGAAAC ATTTTACCCT	3960
	ATGCCATTCT TATCACCACA AACAAATGTGG TACTACAAAT CAGCATTCAA TACTGATCAG	4020
45	ATGAAATTAA TCGATTTAAT TGC GACAATT CAAACGCATA TTGACCAAGG TATCTCAACG	4080
	ATCCTTTATG TTAATTCTGA AATTCTTACA CGTGAGTTAG CAAGATTATA TGTATATGCG	4140
	CACTATAAAG GATTAAAATC ACTTTACTAT ACTAGAAATA AATTATTAAG TGTAGAAGAA	4200
50	TGTACAAGTT GTTCTATCTA ACAATTAAAT GTTGAAAATG ACAAACAGCT AATCATCTGG	4260
	TCTGAATTAG CAGATGATTA GACTGCTATG TCTGTATTTG TCAATTATTG AGTAACATTA	4320

55

	ATGTTTTGGA GACAAAATAT ATCTCAAATG TGGGTTGAAA CAGAAITTTAA AGTATCAAAA	4440
	GACATTGCAA GTTGGAAGAC TTTATCTGAA GCTGAACAAG ACACATTTAA AAAAGCATT	4500
5	GCTGGTTTAA CAGGCTTAGA TACACATCAA GCAGATGATG GCATGCCTTT AGTTATGCTA	4560
	CATACGACTG ACTTAAGGAA AAAAGCAGTT TATTCATTTA TGGCGATGAT GGAGCAAATA	4620
10	CACGCGAAAA GCTATTCACA TATTTTCACA ACACTATTAC CATCTAGTGA AaCAAACCTAC	4680
	CTATTAGATG AATGGGTTTT AGAGGAACCC CATTTAAAT ATAAATCTGA TAAAATTGTT	4740
	GCTAATTATC ACAAACCTTG GGGTAAAGAA GCTTCGATAT ACGACCAATA TATGGCCAGA	4800
15	GTTACGAGTG TATTTTTAGA AACATTCTTA TTCTTCTCAG GTTCTCTATTA TCCACTATAT	4860
	CTTGCTGGTC AAGGGAAAAT GACGACATCA GGTGAAATCA TTCGTAAAT TCTTTTAGAT	4920
	GAATCIATTC ATGGTGTATT TACCGGTTTA GATGCACAGC ATTTACGAAA TGAACCTATCT	4980
20	GAAAGTGAGA AACAAAAAGC AGATCAAGAA ATGTATAAAT TGCTAAATGA CTTGTATTTA	5040
	AATGAAGAGT CATACACAAA AATGTTATAC GATGATCTTG GAATCACTGA AGATGTGCTA	5100
	AACTATGTTA AATATAATGG AAACAAAGCA CTTTCAAACCT TAGGCTTTGa ACCTTATTTT	5160
25	GAGGAACGTG AATTTAACCC AATCATTGAG AATGCCTTAG ATACAACAAC TAAAAACCAT	5220
	GACTTCTTCT CAGTAAAAGG TGATGGTTAT GTATTAGCAT TAAACGTAGA AGCATTACAA	5280
30	GATGATGACT TTGTATTTGA CAACAAATAA CAATTAAATT AAAAGACCTT CACATGTAAA	5340
	GGGAAATAGC GATTCGTTTC GTCTGTCTC CTACATGTTG AAGGTCTTTT TTTATGTGTA	5400
	TCTAACTCAT TATGAGTCTG AGTAAGAAAT CAATGCTCTA AGATGTACAA TGCTATTTAT	5460
35	ATTGGCAGTA GTTGGCGGGG CCCCAACACA GAAGCAGGCG GAAAGTCAGC TAACAATATT	5520
	GTGCAAGTTG GCGGGGCCCC AACATAGAAG CAGGCGGAAA GTCAGCTAAC AATAATGTGC	5580
	AAGTGGCGG GCGCCCAACA TAAAAGCAGG CGGAAAGTCA GCTAACAAATA TTGTGCAAGT	5640
40	TCGGGCGGGG CCCCAACATA AAGAAAAACT TTTTCCTTTA GAAATTATCA CTTCCaCaTG	5700
	AGTTTTACTC ATGTATTCCT ATTTTAAAGT ACACATTAGC TGAGGCTAAT GTTAAGAACC	5760
	ACTACTTAAT CAATCATTAG TAGTTTTTAT CATTTCCACT ATTCCCaGAC ATCaAAATCT	5820
45	TAAGTGTTCT ATTTTACTTT AAGTAAACAA AATACACATT CCGAAAAATT AAATTTTCAGT	5880
	TTAATTGCAA ATATCAATAA AATTGACACT AAATTATTTG AAAGGCTATT GAAATTATGG	5940
	TCAAAAAACG CTACTATTAA TGAGAAATAT TATCAATGAT AATGATTATC ATTAATTTAA	6000
50	AGGGAGAAAA ATTTGTAATG AAGTATTTAT TAAAGGGAAA TATTTTGCTT CTATTACTAA	6060
	TATTGTTGAC AATTATTTTCG TTGTTTCATAG GTGTGAGTGA ACTATCAATT AAAGATTTAC	6120

55

GTATTTTAAT TGCTGGAAGT TCGTTGGCTT TAGCAGGCTT GATAATGCAA CAAATGATGC 6240
 AAAATAAGTT TGTTAGTCCG ACTACAGCTG GAACGATGGA ATGGGCTAAA CTAGGTATTT 6300
 5 TAATTGCTTT ATTGTTCTTT CCAACCGGTC ATATTTTATT AAAACTAGTA TTTGCTGTTA 6360
 TTTGCAGTAT TTGCGGTACG TTTTATTTG TTAAAATCAT TGATTTTATA AAAGTGAAAG 6420
 10 ATGTCATTTT TGTACCGCTT TTAGGAATTA TGATGGGTGG GATTGTTGCA AGTTCACAAC 6480
 CTTCATCTCA TTGCGCACGA ATGCTGTTCA AAGCATTTGGT AACTGGCTTA ACGGGAACCTT 6540
 TGCCATTATC ACAAGTGGAC GCTATGAAAT TTTATATTTA AGTATTCCTC TTTTAGCATT 6600
 15 GACATATCTT TTTGCTAATC ATTTACGAT TGTAGGAATG GGTAAGACT TTAATAATAA 6660
 TTTAGGTTTG AGTTACGAAA AATTAATTAA CATCGCATTG TTTATTACTG CAACTATTAC 6720
 AGCATTGGTA GTGGTGA CTG TTGGAACATT ACCGTTCTTA GGACTAGTAA TACCAAATAT 6780
 20 TATTTCAATT TATCGAGGTG ATCATTTGAA AAATGCTATC CCTCATACGA TGATGTTAGG 6840
 TGCCATCTTT GTATTATTTT CTGATATAGT TGGCAGAATT GTTGTTTATC CATATGAAAT 6900
 AAATATTGGT TTAACAATAG GTGTATTTGG AACAATCATT TTCCTTATCT TGCTTATGAA 6960
 25 AGGTAGGAAA AATTATGCGC aACAATAATA AAAAAATAAT GCTTTTAATT GCAGTAACGT 7020
 TATTAATTAG TATGCTGTAC TTATTTGTAG GTATTGATTT TGAAATATTT GAATATCAAT 7080
 TTTCAAGTCG TTTAAGAAAAG TTCATATTAA TTATTTTAGT AGGTGCTGCC ATTGCAACCTT 7140
 30 CAGTGGTGAT TTTTCAAGCG ATTACAAATA ACCGTCCTATT GACACCATCA ATAATGGGGT 7200
 TAGATGCAGT TTATTTATTT ATCAAAGTAT TGCCAGTCTT TTTATTTGGA ATTCAATCGG 7260
 35 TATGGGTTAC TAATGTATAT TTGAACTTTA TATTAACACT TATAACGATG GTGTTATTCG 7320
 CACTAATCCT ATTCCAAGGT ATCTTTAAAA TCGGACATTT TTCAATTTAT TTTATCTTAC 7380
 TTAATTGGTGT CCTTTTAGGA ACATTTTTTA GAAGCATAAC AGGTTTTATT CAACTGATTA 7440
 40 TGGATCCTGA GTCATTTTTA GCAATACAAA GTAGTATGTT TGCTAATTTT AATGCTTCTA 7500
 ATTCGAATTT AGTTACTTTC TCAGCAGTGC TATTAGTAAT CTTATTAGTC ATTACAATTT 7560
 TACTATTGCC TTATTTAGAT GTATTGCTTT TAGGTCGTGC TGAAGCAATT AATCTTGGA 7620
 45 TATCGTATGA AAAATTAACG CGAATT 7646

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1194 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

5 ATGAATATAT TTnnAAATAA ATTATTATGG ATTGCACCAA TnGCCACTAT GATTATCTTG 60
 GTAATCTTTT CTTTAGCTTT TTATCCTGCA TATAATCCTA AACCAAAAGA TTTACCAATT 120
 GGTATATTAA ACGAGGATAA AGGTACAACG ATTCAAGATA AAAATGTTAA CATTGGTAAA 180
 10 AAATTAGAGG ATAAATTATT AGATAGTGAT TCTAATAAAA TTAAATGGGT TAAGGTTGAT 240
 AGTGAAAAAG ACCTTGAAAA AGATTTGAAA GATCAAAAAA TCTTTGGAGT AGCTATTATT 300
 GATAAAGACT TTTCAAAAGA TGCTATGAGT AAAACACAAA AAGTAGTTAT GGATAGTAAA 360
 15 AAAGAAGAAA TGCAACAAAA AGTTGCTTCA GGTGAAATTC CGCCACAAGT GGTTCACAAA 420
 ATGAAACAAA AAATGGGGAA TCAACAAGTA GAGGTTAAGC AGGCTAAATT TAAAACGATT 480
 GTAAGTGAAG GATCAAGCTT ACAAGGTTCA CAAATTGCAT CAGCTGTGTT AACTGGTATG 540
 20 GGTGATAATA TTAATGCTCA AATTACGAAG CAAAGTTTGG AAACATTAAC GAGTCAAAAT 600
 GTTAAAGTCA ATGCCGCGGA CATCAATGGT TTGACGAATC CAGTAAAAGT GGATAATGAA 660
 AAACTTAATA AAGTTAAAGA TCACCAAGCA GGTGGTAATG CACCATTCTT AATGTTTATG 720
 25 CCAATTTGGA TAGGTTCAAT CGTAACGTCT ATCTTATTGT TCTTTGCATT TAGAACTAGT 780
 AACAAATATCG TCGTGCAACA TCGTATCaTT GcTtCAATTG GACAGATGAT ATTTGCAGTT 840
 GTTGACAGCAT TTGCAGGTAG CTTTGTTTAT ATTTATTTCA TGCAAGGCGT TCAAAGATTT 900
 30 GATTTTGACC ATCCAAATCG TATCGCAATT TTTGTAGCAT TTGCGATTCT TGGTTTCGTG 960
 GGCCTTATTT TAGGTGTTAT GGTATGGCTA GGTATGAAGT CAGTTCCAAT TTTCTTCATT 1020
 TTAATGTCTT TTAGTATGCA ACTTGTAACG TTACCTAAAC AAATGTTGCC TGAAAGTTAT 1080
 35 CAAAAATATG TATATGATTG GAATCCATTC ACACACTATG CAACAAGTGT AAGAGAcTAT 1140
 TATACtTGAA TCATCATATT GAATTAAATA GTACAATGTG GATGTTTATA GGGT 1194

(2) INFORMATION FOR SEQ ID NO: 123:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 558 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

50 GACCGACCTA TACATCCGTA TAAGTATTTT TTGATATAAG TCTTCTAAAT CATAATGATT 60
 AAATCCAAAT GTTTTGATGC GTCGAATAAT TAATGGTTGT AGATCCATTA CTAACtTTTC 120

55

GTATTTCAAA TATTAACTA ACCCCTTCTA TCTAAAATTT AAGGTTAGTT TAATATTGTT 240
 ACATTCAAAA TTTCAAGATG ACGGAAATGT CATTTCTTAT GATGTCCTCT TCGTATTTTT 300
 5 TCAAATTCTG CAAGGATTTT AGAAGATAAC GGAATTCGAG TTCTTGGCTT GTTTTCACTT 360
 ATATCATCTA ATGATTTACT CACATCAATT TCATTTTCTT TTAAATCTCT CCACATTTTCG 420
 CGAGATGATA TTCTATATGC ACCTGATCCA AAGATAGCAT GTTGcTCACT CaTATCACTT 480
 10 GTTACAACCTG TAATATGcTT AGtATGCTTG tCaTAAAGtT CaTAAACCAT AACGGTTCTA 540
 ATGGAAACCA ATCAGCTG 558

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7762 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

25 GCTTCAGACA TnTGATGATA TAATCTCTCA TCATCGATTA ATTCTTTTGC AGCTTGATAC 60
 ACATnTTGCT TATTTGTTCC AATGACTTTT AATGTGCCAG CTTCAACACC TTCAGGACGT 120
 TCTGTAAACAC TTCGCCAAAA CTAAAACTGG CTTATTAAAT GATGGCGCTT CTTCTGAAT 180
 30 TCCACCTGAA TCTGTCAAAA TAAATAAGA TTTTnTAGCA AAATTATGGA AATCTATACG 240
 TCCAAAGGTT CAATCAATTC AATCTGTGCA TGACTACCTA AAATCTTTTG AGCCACCTCT 300
 CGAACTTTTCG GGTTTTTATG CATTGGATAT ACCAGTGCTA AATCAGTATA CTCATCTATT 360
 35 AAGCGTCTAA CCGCTTTAAA TATATTTTCC aTGGGTTTCC CGATATTTTC TCGTCGGTGT 420
 GCTGTCATrA GAATGAATTT kTtGTCTATGG TATTTATCCA TGATGTTAGA TTTATAATTG 480
 TCATCAACTG TATATTTTCAT AGCATCAATC GCAGTATTAC CAGTGACAAC AACACTTTCT 540
 40 GAATATTTCC CTTCACTTAA CAAATGCGAT GCAGCATTTT TAGTAGGTGC AAAATGTAAG 600
 TCAGCTAATA CACCAACTAA TTGTCTATTC ACCTCTTCTG GAAAAGGTGA ATATTTATCA 660
 45 TAACTTCTAA GCCCTGCTTC AACGTGTCCA ATCGGCACCTT GGTTATAAAA TGCCGCTAAA 720
 CCACCTGCAA ATGTCGTCAT CGTATCACCA TGTACAAGTA CCATGTCTGG TTTTCTAAT 780
 TGAATCACTT GTTCTAATTG AGTGATTGAT TTAGAAGTTA TCTCAGAAAG TGTCTGTCCT 840
 50 GATTTTCATA TATTCAAATC GTATTTTGGT TTGATTTCAA AGGTACTTAA TACTGAATCA 900
 AGCATTTCTC TATGCTGTGC TGTAACAACA ACAATTGGCT CGAGCATTTT TTCTTGTTC 960

	ATCTTTTTCA TCAAACACT TATCTCCGAT TCTTCTATTT AGTACCAAAC AATCTATCTC	1080
	CAGCGTCGCC TAACCCTGGT GTGATATATG CTTTGTCAAT aGCTTTTCAT CAAGTGCAGC	1140
5	AATATAAATA TCTACATCTG GATGTGCTTC ATGCATCTTT TCTACGCCTT CTGGTGCTGC	1200
	AATTAAACAC ATGAAGCGAA TATTTTtagc GCCACGTTTC TTCAATGAAG TAATAGCTTC	1260
	AATTGCTGAT GCGCCTGTTG CTAACATAGG ATCAACAACA ATGATTTGTC TTTCAGTAAT	1320
10	ATCTTGAGGT AACTTAGCAA AATACTCTAC AGCCTTTAAT GTTTCGGGAT CTCGATATAA	1380
	ACCGATATGT CCAACTCTGG CTGCAGGTAC TAAACTTAA ATACCATCAG TCATACCTAA	1440
	ACCAGCTCTT AAAATTGGAA CGATAGCTAA TTTTttacca GCTAATCGTT TAGCCGTCAT	1500
15	TTTAGTTACA GCGGTTTCAA TATCAACATC CTGAAGCTCT AAGTCTCTAG TTACTTCATA	1560
	TGCCATCAAC ATACCAACTT CGTCTACAAG TTCTCTAAAT TCTTTAGTAC CTGTATTTAC	1620
20	ATCTCTAATA TAGCTTAGTT TGTGTTGAAT TAATGGATGA TCGAAAACGT GTACTTTACT	1680
	CATAAAAATT ACTCCTATCT TTGTGTATGT TTATTGATAT AGAGGATATT CAGCTGTTAA	1740
	TTTCGCAACG CGTTCCTTAG CTTGTTGTAA TTTTCTTCA TCTTTACTAT TTTTCAATGC	1800
25	TAAACTGATG ATTTTtGCAA CTTCTCtCAA AGCTTTTTCA TCAAATCCAC GCGTTGTTGC	1860
	AGCAGGTGTA CCTAAACGTA TACCACTCGT TACAAAAGGT TTTTCTTGAT CGAACGGAAT	1920
	GGTATTTTTG TTACATGTGA TACCAACTGA ATCTAAAGTC TCTTCAGCTT CTTTACCAGT	1980
30	AAGTCCTATA GACCCTTTTA CATCAACAGC TACTAAGTGA TTATCTGTAC CGCCAGAAAC	2040
	AATTCTAAAT CCTTCATTAA TTAATGCTTC TGCAAGAACT TTTGCGTTTT TAACCACTTG	2100
	TTGTTGATAC GTTTTGAAAT TATTTTCTAA CGCTTCTCCA AAAGCAACTG CTTTtGCTgC	2160
35	AATAACATGC TCAAGAGGTC CACCTTGAAT ACCAGGGAAA ATTGTTTTAT CTATGTCTTT	2220
	TTTATATTCT TCCTTACATA AAATCATACC ACCACGtGGT CCGcGTAATG TTTTGTGTGT	2280
40	TGTAGTTGTT ACAAATCAG CATATTCTAC TGGATTTGGA TGTAACCTG CCGCTACTAA	2340
	TCCTGCAATA TGTGCCATGT CTACCATTAA CTTAGCGTTT ACTTCATCTG CGATTTCTTT	2400
	AAACTTTTTG AAGTCAATTG TTCTTGAATA TGCTGATGCT CCTGCCACAA TAAGCTTAGG	2460
45	CTTATGCTCT AACGCTAATT TACGAACCTC ATCATAATTG ATTCGTTCTG TGTCTTTATC	2520
	TACTCCATAT TCAACGAAAT TGTagAATTT ACCACTAAAA TTAACAGGCG CTCCATGTGT	2580
	CAAGTGACCA CCATGACTCA AATTCATACC TAAAACTGTG TCGCCCATTT CTAATGCAAC	2640
50	TAAGTAAACA GCCATGTTcG CTTGTGAACC TGAATGTGGT TGAACATTGA CATGTTcAGC	2700
	TCCAAACAAT GCTTTAGCAC GATCAATTGC GATGCTTTCA GTAACATCTA CAAACTCACA	2760

55

	TTGTGCTTCC ATAACCGCTT CCGATACAAA ATTTTCCGAT GCGATTAACT CTATGTTGCT	2880
	ATTTTGTCTC TGAAATCTCT TCTCGATTGC TTCTGCGATA ACTTTATCTT GCTTGGTGAT	2940
5	ATAAGACATA AAATCTCCCC TTCTTTCAAA AAAACTTATT GGTATTTAGC ACGTTCGCCA	3000
	CCAATCTTTT TCGGCCTAGA TGTGGCAATA GTTACAATTG CCTGTCCTAC TTGCTTTACT	3060
	GAGGTCCTTA CAGGTACACA TACATGTTTA ATATGCATGC CTATTAACGT TTGACCAATA	3120
10	TCAATTCAC AAGGAACAGT AATATGTTTCG ACCACGATCG GATCCTTCAT ATGCTGAAAA	3180
	GCGTATGTTG CCAAACCTCC TCCAGCATGT ACATCTGGAA CGACGGAAAC TTCTTCCATT	3240
	GTTAATGGAT TATACTGAGA TTTTCTATT GTTATCGCTC TGTTGATATG TTCACATCCT	3300
15	TGAAAGCAA AAGTAACGCC TGTCTCTTTA CTCACAACAT CTAATGCATT AAAAATAGTT	3360
	TCTGCAACTT CCaTCGAACC GACAGTCCCT ATTTTTTCGC CAATGACTTC CGATGTTGAA	3420
20	CATCCAATTA AACATATATC TCCTTTATTA AAAAAGGACA TATCTTTTAA TTCGTCTAAT	3480
	AACATTGTCA AATCTTTTAT AAAAGCCAC CCTTCCTAAA AATAAAAAAG GAATATAGCA	3540
	AAGTGCTACA CTCCTCTATT ATAACTTATT TAACTGTTAA CATATACTAA TTATACAGAA	3600
25	TTCTACTAG CAAATAATAT CTMTTAATTT TAAATTTAA CTTACAAGTT CTTCATAGGT	3660
	ATGTACATAC ATTTCTTTTG TTCCACCGTA TGGATCTATA ACTTCTCCTG CTTCTTTtAC	3720
	ATATTCATGC AATGTGAAAA CATGATTTTG CAAACCAAAG TGTGCCTCTA TTAATTCTTT	3780
30	GTGCGAATAC GACATCGTCA AAATAATATC TGCTTTCAAA TCTGCTTCAG TAAATTGTTG	3840
	CGATAAGGTC GTTTCAGCTA AATGATGTTT TTCAACTAAG TCTTCAACAT AATTCGAAAC	3900
	ACCTTGATTG TTCACAGCGA ATATACCTCT TGATTCAAAT TGATGATTG GCATAACCTC	3960
35	TTTGTCAATA CTTTCCGCTA ATGGGCTACG ACATGTGTTA CCTGTACAAA CGAATAAAAT	4020
	CTTCATAGTT CACATCCTTT AATAATGTGA TTACCTGCAG CTTTTAACAT GCGATTCATA	4080
40	ATTGCTTCTG TATTATCATT CAGCTCAAAG CCGTATATAT ACGCCGCTGA AATATTTTCA	4140
	TTTTCATCAA GTGAATGTAA CACATCATAA AGATTATGAC TTGCTTGTTT AACATCATTG	4200
	TCATCCTGAC ATAATTGAAT GAATTGCGCT TCACTTGGTA TAAACGCCAC CTTATTACTC	4260
45	GGCACAATAA AAGCTATAGA AGACCAATCT TTACCGTCAT TTCCAATTTT GCTCTCAATA	4320
	TCTGTAATAA TTGTAAGTGG TGTATTGGGT GAGTAATGCT TATACTTCAT ACCTGGTGCA	4380
	ATTGGCTGTT CAGTATCATT ATAATCAGCA TGGGCGATAC TATTGGAAG TATTTCTGTA	4440
50	ATCATTGCTG CTGTTATAGA ACCAGGTCTT GCAATTTTAT AAGGAAAAGA TGTGCAATCT	4500
	AAAACCGTAC TTTCTAATCC TTCTTCACTT TGTTTCAGCTT GAACAATACC ATCGATACGG	4560

55

EP 0 786 519 A2

	GCACTTGGAG CAGCTAGAGG TTCATTTATG ATTTGTAATA ATTGTCTACC TACAGAATGG	4680
	CTTGGCATT C TAACAGCAAC TGATGATAAA CCTCCAGAAA CTTTTCGACA TAGATAGCCT	4740
5	AGCTTTAACG GCAATATAAA CGAAATAGGG CCCGGCCAGA ATGCCTGCAT TAACTTTTCT	4800
	ACGCGTGGAT CCAAAGTATA TGTAAATCT TTTAATTGAC CTTTACTGTG TATATGAACA	4860
	ATAAGCGGAT TGTGAGATGG ACGGCCTTTA GCTTCATATA TTTTAGCTAC AGCTTCTTCA	4920
10	TCTGTCGCAT TTGCTGCAAG TCCATAAACT GTTTCAGTTG GTAAACCTAT TAAACCACCG	4980
	TTTAAACAA TGTCTTTTAT TTCATTAATT TTAGGATATT GCTGTAAATC TTCATTATAT	5040
	TCTCTAACAT CCCAAATTTT AGTATCCAAC TTAATCACGC CTTTCTTATT TATCATAATA	5100
15	TAAAGCAAAA AGCTATGCAC TTAAC TAATC ATAGCAAAGG CATAACTTCT AATTACCATT	5160
	TAAATGAGAC GATTCGATCG TGGCCATTTA TATCTTTAAT AATGTCGATT TTTTGTGAG	5220
	GAAATTTATT TAAAATTATT GATTTAAGTG CCTCACCTTG ATTGTAACCA ATTTCAAAAA	5280
20	CAACTGGGCT GCCTTTTTC ATAACGTGAG GTAAATCTTC AATGATTGAT TCATAAATAG	5340
	CATATCCATG GTTATCTGCA AACAATGCCT GATGTGGTTC GAATCTCGTA ACCGTTGGAG	5400
	ACATCGTAAC CATATCTTTT TCATCTATAT ATGGTGGATT AGATATCAAG CCGTTCAACT	5460
25	TGATACCTTC ATTAATTAAG GGCTTTAATG CATCCCCTGT TAAAAATTGT ATTTGTGATT	5520
	GATGCTTCTC AGCATTATTA CGAGCCATAT TCATTGCTTC AAGTGAAATA TCAGTAGCAA	5580
30	TAACATTTAA ATCCGGCTTT TCACATTTCA AAGTAATTGC AAGTACACCA CTACCCGTTT	5640
	CGATATCTAC GATTGTTGCA TCATCTTCTA ACTGTTGTAA GAAATGCAAC ATTACTTCTT	5700
	CAGTTTCAGG TCTTGGTATC AAACAATTTG AGTTTACATC AAACGTTCTA CCATAAAATG	5760
35	AGGCAAAGCC AACTATATAC TGTATAGGCT CTCCTAATAA CATACGTTGT AATGCTAAGT	5820
	CGAATCTCAT AATCATCGCT TTCGGCATAT CATCATGCAT GTGGACTACA AAGTCCGTAC	5880
	GCGTCCATTG AAATACATCT AACATTAACC ATTCAGCTCG TGTTTGTTCA AACCCTTTTT	5940
40	GTTGTGTTAA ATGAATTGCT TCATCTAACT TTTCTTTATA ATTCACCATT ATTAAGTTCT	6000
	TTCAATTTAT CTGTCTGCTC TGATAAAGTC AGTGCATCTA TAATTTCTTC TAAATGGCCT	6060
	TCCATAATTT GCCCTAATTT TTGAAGCGTT AGACCTATAC GATGGTCTGT TACACGGCTT	6120
45	TGTGGATAAT TATAAGTTCT AATACGTTCT GAACGATCAC CAGTACCGAC TGCTGATTTA	6180
	CGTTGTGACG CATACTTTTG TTGTTCTTCT TGAAC TTTCATCGTATAA ACGTGCTTTT	6240
50	AACACTTTCA TTGCTTTTTT ACGGTTTTGA ATTTGAGACT TCTCAGAAGA TGTTGCAATG	6300
	ACACCAGTTG GTAAATGGGT AATACGTA CTGAGATCAG TTGTGTTTAC GTGCTGACCA	6360

55

ACATCTTCAA CTTCTGGTAA AACTGCCACT GTAGCTGTTG AAGTATGAAT ACGTCCACCT 6480
 GATTCTGTTT CAGGCACACG TTGAACGCGG TGCGCACCAT TTTCAAATTT CAATTTACTA 6540
 5 TACGCGCCAT TACCAGAAAC TGAGAAACTA ATTTCTTTGT AACCACCATG GTCACCTTCA 6600
 GACGCTTCTA CTATTTTCAGT TTTGAATCCT TGTGATTCAG CATACTTTGA ATACATACGC 6660
 ATTAAATCAC CAGCAAAAAT CGCAGCCTCA TCACCACCTG CTGCTGCTCT TATTTCTACA 6720
 10 ATAACGTCTT TGTCATCATT AGGATCTTTA GGAATCAATA ATATTTTAAG CTCTTCTTCA 6780
 AGATTTGGAA GTTCAGCTTT AATACCATTA CTCTCCTCTT TTAACATTTT TACTTCTTCT 6840
 TTATCATCAG TCTCACTTAA CATTTCTTCA ATATCAGCTA ATTCTTCTTT TTTAGCTTTA 6900
 15 TAGTTACGAT AAACATCTAC AGTTTTTTGT AAATCAGCTT GCTCTTTAGA ATATTACGT 6960
 AATTTATCTG AATCATTTAC AACATCTGGG TCACCTAACA GTTCATTTAA CTGTTTCGTAT 7020
 CTTTCTTCTA CAATATCTAA TTGATCAAAC ACTTATAATT CCTCCTTATT ATTATCACTA 7080
 GGTGCTACGA TATGGTGCGC GCGACAACGT GGCTCATAAC TTTCATTGGC ACCTACTAAG 7140
 ATAATCGGAT CATCGATTTT AGCTGGTTTA CCATTTATTA ATCGTTGCGT TCTACTAGAT 7200
 25 GAAGAACCAC AAACAGCACA AACTGCTTGA AGTTTCGTTA CTTGTTCACT GACAGCCATC 7260
 AATTTAGGCA TTGGTTTCGAA CGGTTTCGCCC CTAAAATCCA TATCTAATCC AGCAACAATA 7320
 ACACGGTGTC CATCTGCTGA TAGTTTTTCT ACTATACTTA CAATTTCATC GTCAAAAAAT 7380
 30 TGCACTTCGT CTATTCCTAT AACATCAACA TTAGTTAAGT CGTGCGTCAT AATTTCACTT 7440
 GCTTTAGAAA TATTAATCGC TTCAATGGCA TTACCATTAT GAGAGACCAC TTTTCTTTA 7500
 TGATATCGAT CATCAATCGC CGGTTTAAAT ACAACGACTT TTTGTTTAGC GTATATACCC 7560
 35 CTTCTTAGAC GTCTTATTAG TTCTTCGGAT TTACCGCTAA ACATACTACC TGTAATACAT 7620
 TCTATCCAAC CGGAATGGTA AGTTTCATAC ATTGAGAGTn CCACCTTTTT CAAAACATAA 7680
 TCGCTTTATT ATATCATATT TCAAATATTC ATAAATGTCT TTTTCATAAT TATATCGATA 7740
 40 TTGTACATGA ACAATTATTT TA 7762

(2) INFORMATION FOR SEQ ID NO: 125:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2583 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

55

	TAAAAAAATT ATTATCAATG ATGAAC TAGA ATTGACTGAA TTCCACCAAG AACTTACTTA	120
	TATTTTAGAC AACATAnAAG GGAATAATAA TTATGGTAAG GAATTTGTTG CAACCGTTGA	180
5	AGAAACATTC GACATTGAAT AaAGCGGGGT GgaAGCACTA TGAATCAATG GGATCAGTTC	240
	TTAACACCTT ATAAGCAAGC GGTTGATGAG TTGAAAGkGA AcTTaAAGGC ATGCGCAAAC	300
	AATATGAAGT TGGTGAACAA GCGTCGCCAA TAGAATTGT TACTGGTCGT GTTAAACCAA	360
10	TCGCTAGTAT TATAGATAAG GCAAACAAAC GACAAATACC ATTTGATAGG TTAAGAGAAG	420
	AAATGTACGA TATCGCTGGT TTAAGAATGA TGTGCCAATT TGTGAAGAT ATTGATGTTG	480
	TCGTCAATAT TTTAAGACAA AGAmAAGATT TTAAAGTAAT TGAAGAACGA GATTATATTC	540
15	GTAACACTAA AGAAAGTGGT TACCGCTCGT ATCATGTCAT TATTGAATAT CCAATTGAAA	600
	CATTACAAGG CCAAAAATTT ATATTGGCTG AGATTGAGAT TCGTACATTA GCAATGAATT	660
20	TCTGGGCAAC GATTGAACAT ACTTTACGAT ATAAATATGA TGGTGCTTAT CCGGATGAAA	720
	TTCAACATCG TTTGGAAAGA GCGGCAGAAG CAGCGTATTT ACTTGATGAA GAGATGTCTG	780
	AAATTAAAGA TGAAATTCAG GAAGCTCAAA AATATTACAC GCAAAAACGT TCTAAAAAAC	840
25	ATGAAAATGA TTAACGAGGT GTTATAAATC ATGCGTTATA CAATTTTAAC TAAAGGTGAC	900
	TCCAAGTCTA ATGCCTTAAA GCATAAAATG ATGAACTATA TGAAAGrTTT TcGCATGaTT	960
	GaGGATrGTG AAAaTCCTGA AATTGTTATT yCAGTTGGTG GTGACGGTAC ATTACTACAA	1020
30	GCATTCCATC AGTATAGCCA CATGTTATCA AAAGTGGCAT TTGTTGGAGT TCATACAGGT	1080
	CATTTAGGAT TTTATGCGGA TTGGTTACCT CATGAAGTTG AAAAATTAAT CATCGAAATT	1140
	AATAATTCAG AGTTTCAGGT CATTGAATAT CCATTGCTTG AAATTATTAT GAGATACAAC	1200
35	GACAACGGCT ATGAAACAAG GTATTAGCA TTAAATGAAG CAACGATGAA AACTGAAAAT	1260
	GGCfCAACAC TTGTTGTGGA TGTAACTTA AGAGGGAAAC ACTTTGAGCG ATTTAGAGGC	1320
40	GATgGATTAT GTGTATCAAC ACCTTCGGGT TCAACGGCTT ATAACAAAGC GCTAGGTGGC	1380
	GCACTGATAC ATCCTTCACT TGAAGCAATG CAAATTACAG AAATTGCCTC GATAAATAAT	1440
	CGTGTGTTTA GAACGGTAGG ATCACCACCT GTATTACCAA AGCATCATAC ATGTTTAATA	1500
45	TCACCAGTTA ATCATGATAC CATTAGAATG ACGATAGATC ATGTTAGTAT CAAACATAAA	1560
	AATGTTAATT CAATACAATA CCGTGTAGCA AATGAAAAAG TGAGGTTTGC ACGTTTTAGA	1620
	CCATTCCCAT TCTGGAAACG TGTGCACGAT TCTTTCATAT CAAGTGATGA AGAACGATGA	1680
50	AATTTAAGTA TCATATATCA CAACAAGAAA CTGTtAAAC TTTTTAGCA CGACATGATT	1740
	TTTCTAAGAA GACAGTGAGC GCCATTAAAA ATAATGGCGC TTTAATTGTT AATGATGAAC	1800

55

AAATACCGAG TGTAAATTTA ATACCTTATG CTCGTAAGCT AGAAGTATTG TATGAAGATG 1920
 CTTTTATCAT CATAGTTACT AAACCAAACA ATCAAAATTG TACGCCTTCG AGAGAACATC 1980
 5 CTCATGAAAG TTTAATCGAA CAAGTACTAT ATCATTGTCA GGAACATGGT GAAAATATTA 2040
 ACCCACATAT TGTTACGCGT CTAGATCGTA ATACAACCTGG TATTGTGATA TTCGCTAAAT 2100
 ATGGACATAT CCATCATTTA TTTTCTAAAG TAAACTTGAA AAAAATATAT ACTTGCCTTG 2160
 10 TATATGGTAA AACC CATACA TCTGGTATTA TTGAAGCTAA TATTAGACGG TCAAAGGATA 2220
 GGATTATAAC TAGAGAAGTT GCCTCGGATG GTAAATACGC TAAAACATCT TATGAAGTAA 2280
 TAAATCAGAA TGATAAATAC AGTTTATGCA AAGTTCATTT GCATACGGGA CGTACACATC 2340
 15 AAATTCGTGT ACATTTTCAA CATATTGGGC ATCCAATTGT GGGAGATTCT TTGTATGATG 2400
 GTTTTCATGA CAAAATTCAT GGTCAAGTAC TGCAATGTAC GCAAATATAT TTTGTTTCATC 2460
 20 CAATCAATAA GAACAATATT TATATTACAA TTGATTATAA GCAATTACTT AAATTATnCA 2520
 ATCAACTCTA ATnCACACAG GGGGTGTAAG TATGTCAATG AnCACAGATG AAAAAGAGCG 2580
 TGT 2583

25 (2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1818 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:
 35 ATCAAGTGAT ACATTTAACT GTTAAAGGAT TAAAnAGATGC TCAAGTTAAA AAATCnGGAT 60
 ATATACAATA TGAATTTGTT AAAGAGGATT TnACAGATTT ATTnGCAATT ACGGATACAG 120
 TAATAAGTAG AGCTGGATCA AATGCGATTT ATGAGTTCTT AACATTACGT ATACCAATGT 180
 40 TATTAGTACC ATTAGGTTTA GATCAATCCC GAGGCGACCA AATTGACAAT GCAAATCATT 240
 TTGCTGATAA AGGATATGCT AAAGCGATTG ATGAAGAACA ATTAACAGCA CAAATTTTAT 300
 TACAAGAACT AAATGAAATG GAACAGGAAA GAACTCGAAT TATCAATAAT ATGAAATCGT 360
 45 ATGAACAAAG TTATACGAAA GAAGCTTTAT TTGATAAGAT GATTAAAGAC GCATTGAATT 420
 AATGGGGGGT AATGCTTTAT GAGTCAATGG AAACGTATCT CTTTGCTCAT CGTTTTTACA 480
 50 TTGGTTTTTG GAATTATCGC GTTTTTCCAC GAATCAAGAC TTGGGAAATG GATTGATAAT 540
 GAAGTTTATG AGTTTGTATA TTCATCAGAG AGCTTTATTA CGACATCTAT CATGCTTGGG 600

55

	CTCATGTTAA AGCGCCACAA AATTGAAGCA TTATTTTTTG CATTAAACAAT GGCATTATCT	720
	GGAATTTTGA ATCCAGCATT AAAAAATATA TTCGATAGAG AAAGACCTAC ATTGCTGCGT	780
5	TTAATTGATA TAACAGGATT TAGTTTTCTT AGCGGTCATG CTATGGGATC AACTGCATAT	840
	TTTGGAAGTG GTATCTATCT ATTAAATCGA TTAAATCAAG GTAATTCAA AGGTATTCTT	900
	ATAGGGTTAT GTGCAGCTAT GATTTTATTG ATTTCCATAT CACGTGTATA TCTAGGTGTA	960
10	CATTATCCAA CAGATATTAT TGCCGGCATT ATTGGTGGAT TATTTTGcAT TATTTTATCA	1020
	ACGTTATTAC TTAGAAATAA ATTAATAAAT TAAATAGTAA AAAACAAAA GCAGTAAACC	1080
15	TAAAGTGTCG TAAGGGTTTA CTGCTTTTAT AAAACGTTGT TATAACGTAT ATTGTCTTTT	1140
	ACGGGCATAT AAnAGGGGAA TATTTGAnAA TGACCAATCC AACAAGAACG AAACGTTGTG	1200
	GGGGGGATGT TCTATGTGGT ATTGATAATC ATTTTCAACT ACTATTATAC ATTAGTGAGA	1260
20	ATCATTGTCA ATTAGAAACT AAACTTTTTT TTGAATATTT TTTAAGAATA GTAAATAAAA	1320
	CGCATGATTA CGCTATTTTA GAAAAATAAA AAATTTGTAT TTCTCATTAG AATTAGAATA	1380
	TTTAAAAGTG ATGAGGTTTA AACATTATAT TGTTTACATA CTCCTTTTGA ATTCATACAT	1440
25	TATGAAATGT tACTTCCAAG TTCAAAATCG CACATTGAAA TGATGTGTGA AATGTTTAAA	1500
	CTACGGTCAT tTTGTGmAAA TAAAGrTAAT AACTATTCAT TTTACAATAG TGAAAAGTCA	1560
	GIATATGACA ACAATTAATA TTGCGGTAAG GCCTTGTGTT ACAGTATTCT ATATTTAAGT	1620
30	ACTGCAATCA GAATTAACAG AATGCCATTA ACTGATTATT AAATATTTGA GTTAATAAAT	1680
	AATTAATGAT TGTAGCTTGA AAAATTTTAA ACATGGTTAT TGATTTGTGA TAAAATTTAA	1740
	ACGTAAACAA ACTAATTTAA AAAGCAACTA TTGTATAGAA AAATACAAAA TTTAAATAT	1800
35	ATTACCTTAT TAGAAAAA	1818

(2) INFORMATION FOR SEQ ID NO: 127:

40	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 12658 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

	TGTTTAAACA ATAGGGGGAA TCTTATGATT GAAAAATTAG TAACCTTTTT AAATGAGGTT	60
50	GTTTGGAGTA AGCCATTAGT TTATGGTTTG CTAATTACTG GTGTGCTATT TACATTGCGT	120
	ATgCGATTTT TTCAAGTTAG ACATTTTAAA GAAATGATTC GATTAATGTT TCAAGGAGAG	180

55

GGTACAGGTA ATATTGTCGG TGTATCTACT GCAATATTTA TAGGAGGACC TGGTGCAAGTA 300
 TTTTGGATGT GGATTACTGC GTTTTtagGT GCAAGTAGTG CTTTTATTGA ATCTACACTT 360
 5 GGTCAAATAT TCAAGAGAGT TGAAAATAAT GAATACCGTG GTGGACCAGC GTATTATATT 420
 GAATATGGTA TTGGTGGTAA ATTTGGTAAA ATTTACGGAA TTATCTTTGC TATTGTTACG 480
 ATTATCTCAG TAGGTCTATT GCTTCCTGGT GTGCAATCTA ACGCTATAGC AAGTTCTATG 540
 10 CATAATGCGA TTCATGTTCC ACAATGGTTA ATGGGTGGTA TTGTTGTAGT TATTTTGGGA 600
 TTAATTATTT TTGGTGGTGT ACGTATTATT GCCAATGTTG CAACAGCCGT TGTACCATT 660
 ATGGCAATTA TTTACATACT GATGGCTGTC ATTATCATT 720
 15 CCAGCGTTAT TTGCATTAAT TTTCAAATCA GCATTGGAT TACAATCTGC TTTTGGTGGT 780
 ATCGTTGGCG CAATGATAGA GATTGGTGT AAACGTGGAT TATATTCAA TGAGGCTGGT 840
 20 CAAGGTACAG GTCCACACGC AGCAGCGGCa gcAGaAGTAT CACATCCAAG TAAACAAGGT 900
 CTAGTACAAG CATTTTCAGT TTATATTGAT ACATTATTTG TATGTACTGC AACTGCTCTG 960
 ATTATACTTA TTTCTGGTAC ATATAATGTG ACTGATGGTA CGGTTAATGC GAATGGCACA 1020
 25 CCGCATTTAA TTAAAGATGG CGGTATTTAT GTTgAAAATG CAACAGGTAA AGATTATTCA 1080
 GGTACTGCGA TGTATGCACA AGCCGGCATt GATAAAGCGT TCCATGGCAG TGGTTATCAA 1140
 TTTGATCCTA CTTTCTCTGG CGTAGgTTCG TACTTTATTG cATTTGCTTT ATTCTTCTTT 1200
 30 GCATTTACTA CAATTTTGTC GTACTACTAC ATTACAGAAA CAAATGTTGC TTATTTAACG 1260
 CGTAATCAA ATAATCAAGT TTCATCGATA TTTATTAATA TTGCTCGTGT GATTATTTTG 1320
 TTCGCTACAT TTTACGGTGC AGTTAAAACA GCTGATGTAG CATGGGCATT CGGTGATTTA 1380
 35 GGTGTAGGTC TAATGGCTTG GTTAAATATC ATTGCGATT GGATTTTACA TAAGCCTGCC 1440
 GTAATGCTT TAAAAGATTA TGAAATTCaa AAGAAACGTT TAGGCAACGG TTATAATGCA 1500
 GTTTATCAAC CTGATCCGAA TAAATTACCT AATGCTGTCT TTTGGTTGAA GACGTATCCA 1560
 40 GAACGTTTAA AACAAAGCAG TGCCAAAAAG TAATCTACTT TTGTTTATAG TATATGTAGT 1620
 GATCATTTGA TAAAAAAGAA AAGTATTGAG AATTTTAGGt GCTCAGAAAT TTGAATTTTA 1680
 45 AAAATATAGT GTCTCTTGGT ACAATAACAA TACAACACT AGGGGCACTT TTTTATGTCA 1740
 GAATTTAAAA CTGGTAAGAT TAATAACAT GTTTTATATA GTAATATTTT AAATAGAGAT 1800
 GTCACGTTAA GTATTTATTT ACCAGAATCT TATAATCAAC TTGTTAAATA TAATGTCATT 1860
 50 CTTTGCTTTG ACGGATTAGA TTTTTTACGT TTCGGGAGAA TACAACGTAC ATATGAATCG 1920
 TTAATCAAAG AAGCGCGTAT TGATGATGCG ATCATTGTTG GATTCCATTA TGAAGACGTT 1980

55

	GTCGGTAAAG	AAATATTGCC	ATTTATTGAC	TCGACGTTTT	CTACACTGAA	AGTAGGTAAT	2100
	GCAAGGTTAT	TAGTAGGGGA	TAGTTTAGCG	GGTAGTATTG	CCTTATTAAC	GGCGTTGACC	2160
5	TATCCAACGA	TTTTTAGTCG	TGTAGCAATG	TTAAGTCCAC	ATTCAGATGA	AAAAGTATTA	2220
	GATAAGCTAA	ATCAATGTGC	AAATAAAGAA	CAATTGACAA	TTTGGCATGT	CATTGGTCTA	2280
	GATGAAAAAG	ATTTTACTTT	ACCAACAAAT	GGTAAGCGTG	CCGATTTCTT	AACACCGAAT	2340
10	AGAGAATTAG	CTGAACAAAT	TAAGAAATAT	AATATAACTT	ATTATTACGA	TGAATTTGAT	2400
	GGTGGTCACC	AATGGAAAGA	TTGGAAACCA	TTGCTGTCAG	ATATATTATT	GTATTTTTTA	2460
15	AGTAAAAACA	CAGATGATCA	ACTTTATGAA	TAATTTACAT	TAGTAGATTT	AGTATGAATT	2520
	GTCTTCATAT	AGTCTGGTCT	ATAATATAAT	TTATAAAAGA	TTTTACTGTT	TAATTTAATT	2580
	TAAATTTGAC	GAAATTGCAA	AAGATGTATA	ATGAATTATT	TTTAATGTAA	CGGTTTTCAA	2640
20	AGAAATTTGA	TATAATAGCA	ATAGGTTAAA	CAAAGGAGGA	ATTCAGATGA	TTTTAGGATT	2700
	AGCATTAATT	CCATCAAAGT	CATTTCAGA	AGCGGTGGAT	TCTTACCGTA	AAAGATATGA	2760
	TAAACAGTAT	TCACGAATTA	AACCACATGT	GACAATTAAA	GCGCCATTG	AAATTAAAGA	2820
25	TGGTGATTTA	GATTCTGTCA	TTGAACAGGT	TAGAGCTCGT	ATTAATGGTA	TACCAGCAGT	2880
	AGAAGTTCAT	GCTACAAAAG	CTTCTAGCTT	CAAACCAACG	AACAATGTGA	TTTACTTTAA	2940
	AGTTGCGAAG	ACGGACGACT	TAGAAGAATT	GTTTAATCGC	TTTAATGGAG	AAGATTTCTA	3000
30	TGGAGAAGCT	GAACATGTTT	TTGTGCCACA	CTTTACAATA	GCACAAGGAC	TATCTAGCCA	3060
	AGAATTCGAA	GATATTTTTG	GTCaAGTAGC	ATTAGCTGGG	GTAGACCcTA	AAGAAATTAT	3120
	CGATGAATTA	ACTTTGTTAC	GTTTIGACGA	TGACGAAGAT	AAATGGAAAG	TTATTGAAAC	3180
35	GTTTAAATTA	GCTTAAGTAA	CATAATAGTA	TTGTTAATCG	TAGTATGTTT	GAATTAATAA	3240
	GAAAATGGTC	ATTTTATTG	AATGTAATAA	AAATGACCAT	TTTCTTTATT	TTAAATACG	3300
	TTTTAACCTT	ACTTAGCTTT	TTCTCTATTT	ACTATAAAGT	TCCTTCCATA	AAATACAGCT	3360
40	AAGACTAAAA	AGATTAATGC	CGAGAAATAA	AATGTATTGT	TTAAATTGTT	GGTAAATTGT	3420
	GTAATTAATC	CGCCAAATAA	TGGCCCTATC	ATTGAGCCGA	ATCCTTGGAT	ACTATTAAAA	3480
45	ACACCCCAAG	TTTCTTCTTG	TTCATCTGAT	TTGATAAATC	GTGCCATAAA	GGTATTCCAT	3540
	GCTGGTAATA	AGATGCCATA	CATTAGACCG	ATAGCTAAAG	CGATAATCCA	CAAGATGTGA	3600
	ATATTAACAA	TCATAGATAG	AGTAAAAATT	AATATCATGT	ATAAAATAAA	TCCGCTTAGA	3660
50	ATAACACCAT	ACATAAAGTT	TCTGCTGCGG	TTATCTATTA	GTTTCGATAA	AAATAGCATC	3720
	GAAACTGCAC	AGCCGATACC	ACCAATAATG	ATTGCAACAG	TATATTCAAT	TGTGCTTACG	3780

55

	TGTAAAAGAA TACCAGGGAA CaACAATAAA TGGcGCTTTG TCACATCAAC AATTGTCTC	3900
	AATTGAGCTT TAACTGGACG AGTATTATAA TTTGTAACT TTACATCGAC AAAATAATAT	3960
5	AATATCCATG CAATTAAAC GACTAAAGAC ATCATGAAGG CAAAGCGTGT TGGGTGCACT	4020
	TTGATAAGTA GATTCATAAA AACCATACCT ACCAATAGGC CTAACAACCA TGAAAAATAA	4080
	ACATAGCCCA TTTGTTTGCC ACGTTTATCT TCTTCAACAC TGGATAACAT AATGACCCAA	4140
10	ATAGGACTAA CTGCAATACC GAGCATCATA GCACTAAATA TGATTACAAA AGGTGATGCT	4200
	GGAAACCAAA TAACTAAAAA TAACTTGTA AATGCTAAAA TAAATCCAGT CGTTAAAACG	4260
	ATTTTGTGTC CGAATTTT CAGTAAAAAT CCTATAACAA AGTTTGTAGA TGCATCAGCA	4320
15	ATAAAATGTA TTGAAATGC TAGAGACGTT ATTGCTACAG CAATGGATGT AACTGTTGGC	4380
	AAGAAATTAA TATAGCTTAG GATATACATG CCTCTCGCAA ATTCCATTAA AAATAAGATA	4440
20	ATAAGCaTTA AAATGAAATT TTTATGATTA GCGTAATTAT TTAACGAAGA ATCTGCATA	4500
	TAAAGGAACC TTTCCATAAA TCTCTGTGG TTGTGATGAA TGACCGATTA AATCAAGTAA	4560
	GTCTCGACAT ATTGTCTGTG TAGCATACTT AATTTTATCT TGTTCCATTG TACTAATCAT	4620
25	GTTAGTTAAT TGCTCATTAC CGTAGTTAA ACTTGCTACA ATTTTATTG CTTCTTCTGG	4680
	AGTATCAGCG ATTTTACCAA AACCTTTTTC TTCAAAGTAA AGGGCATTTC CAAGCTCTTG	4740
	ACCAGGTGCA GGATTAGGA AAATCATTGG AATACAACGG GCGAAACCTT CAGTTATTGT	4800
30	GATACCACCA GGTTCGTAA TCATAAGTTG ACTTGATGCC ATCCATTCTAT TCATGTGTTT	4860
	GGTATAACCT AGAATCAATA CATTCTCGTT AGATTTAAAC TTAGCTGTTA AAGAACGCTT	4920
	TAGCTCTTTG CTCTTACCAC AAATCATAAC TACTTGTGCA TTTGCaTTT tCGCTAATAT	4980
35	ATCAGTAATC ATCGTGTC AAACCTTTAGA TACACCAAAT GCACCAGCTG aCATTAAAAT	5040
	AGTTTGCTTA TCTGGATCTA AGTTGTTGTC TATTAACCAC TGCTTTTGAT TAATAGGCGT	5100
	TTCAAATTG TTATCAATAG GAATACCTGT CaCTTTAACT GTTGAAGGAT CAATACCTAC	5160
40	GTCTATGAAG TCTTGTTTCG TTTCTTTTGT TGCCACATAA TATCTTGTTG AATACGGCGT	5220
	AATCCAGTTT TTATGTAAGC GATAGTCTGT CATCACTGTA GCAACTGGAA TATTAATGTT	5280
45	AAATTGCTCA GTTAGTACCG ACATAACTGG TGTAGGAAAC GTTAATAATA TTAAATCTGG	5340
	CTTTTCTTTT ATCAATAAAT TAATTAACCT ATTAAGTCCA TAGTATTTGT AAAACATTT	5400
	GTCTAGTTTA TCTGGGCGGC TGTAATAAAA CCCTTTGTAC ATATTTCTAA AATATTTAAA	5460
50	GCTATTGATA TACCATTTT TACAAATAGA AGTCAAAATT GGATGAGCTT CCATAAATAA	5520
	ATCGTGCTCA ATGACGCTTA AATGGTCTAG ATTCATATCA TTAAGTTGAT TAACGATACT	5580

55

	TTGAGTAACC ATTAATAGCC ACCCTCCGTT AGTTTGAAAA TTTTATTTAA GTGTAACCTTA	5700
	TTTTACGGCA TTATAAAAGA AATAAAGACG CAAAGTCGTT ACATTTATAG CAATTTTAAT	5760
5	CTATAGATGA ATTGATACAA AATAAAACGT TATTTTATAA AGCAATTTAT TGTTCCTATGT	5820
	TTTATTTGTA TATTTAAAT TATCCAGTAT ACAATTATAG CATATTTTGT GAAACAATTA	5880
	TGATATTATA CCATGTTACA AGATGGTTTT AATAATTTAA GATGAGCCAT AATTGTAAAA	5940
10	CTAATTCATA ATACCGTATG TTTTATTTTT AATAGTAGAA ATTAGAAAAT GCTGATTAGT	6000
	AGGATATAAC AGTGAAATTA TAAATTTATT AACATCAACA AAACGTGTAT AATAAACATA	6060
15	TTGTAGAAAA AGGAGCGGTT CAGTTTGGAT GCAAGTACGT TGTTTAAGAA AGTAAAAGTA	6120
	AAGCGTGTAT TGGGTTCTTT AGAACAACAA ATAGATGATA TCACTACTGA TTCACGTACA	6180
	GCGAGAGAAG GTAGCATTTT TGTCGCTTCA GTTGGATATA CTGTAGACAG TCATAAGTTC	6240
20	TGTCAAAATG TAGCTGATCA AGGGTGTAAG TTGGTAGTGG TCAATAAAGA ACAATCATT	6300
	CCAGCTAACG TAACACAAGT GGTTGTGCCG GACACATTAA GAGTAGCTAG TATTCTAGCA	6360
	CACACATTAT ATGATTATCC GAGTCATCAG TTAGTGACAT TTGGTGTAaC GGGTACAAAT	6420
25	GGTAAACTT CTATTGCGAC GATGATTCAT TTAATTCAAA GAAAGTTACA AAAAAATAGT	6480
	GCAATTTTAG GAACTAATGG TTTCCAAATT AATGAAACAA AGACAAAAGG TGCAAATACG	6540
	ACACCAGAAA CAGTTTCTTT AACTAAGAAA ATTAAAGAAG CAGTTGATGC AGGCGCTGAA	6600
30	TCTATGACAT TAGAAGTATC AAGCCATGGC TTAGTATTAG GACGACTGCG AGGCGTTGAA	6660
	TTTGACGTTG CAATATTTTC AAATTTAACA CAAGACCATT TAGATTTTCA TGGCACAATG	6720
	GAAGCATACG GACACGCGAA GTCTTTATTG TTTAGTCAAT TAGGTGAAGA TTTGTCGAAA	6780
35	GAAAAGTATG TCGTGTTAAA CAATGACGAT TCATTTTCTG AGTATTTAAG AACAGTGACG	6840
	CCTTATGAAG TATTTAGTTA TGGAAATGAT GAGGAAGCCC AATTTATGGC TAAAAATATT	6900
	CAAGAATCTT TACAAGGTGT CAGCTTTGAT TTTGTAACGC CTTTGGGAAC TTACCCAGTA	6960
40	AAATCGCCTT ATGTTGGTAA GTTTAATATT TCTAATATTA TGGCGGCAAT GATTGCGGTG	7020
	TGGAGTAAAG GTACATCTTT AGAAACGATT ATTAAAGCTG TTGAAAATTT AGAACCTGTT	7080
45	GAAGGGCGAT TAGAAGTTTT AGATCCTTCG TTACCTATTG ATTTAATTAT CGATTATGCA	7140
	CATACAGCTG ATGGTATGAA CAAATTAATC GATGCAGTAC AGCCTTTTGT AAAGCAAAAG	7200
	TTGATATTTT TAGTTGGTAT GGCAGGCGAA CGTGATTTAA CTAAAACGCC TGAAATGGGG	7260
50	CGAGTTGCCT GTCGTGCAGA TTATGTCATT TTCACACCGG ATAATCCGGC AAATGATGAC	7320
	CCGAAAATGT TAACGGCAGA ATTAGCCAAA GGTGCAACAC ATCAAAACTA TATTGAATTT	7380

55

	GTTTTAGCAT CAAAAGGAAG AGAACCATAT CAAATCATGC CAGGGCATAT TAAGGTGCCA	7500
	CATCGAGATG ATTTAATTGG CCTTGAAGCA GCTTACAAAA AGTTCGGTGG TGGCCCTGTT	7560
5	GATTAATAAA AGATTTATTG ATGAAGGTAA AACTATTGAT GTTTATTTAT TCGAAGCATT	7620
	AAATAACCAG ATAATCATTG CTATACCAGA TTGGTTTTGG TCATATCAGA TGGCAATGAC	7680
	ATTAGATGAA GAAACTTGTT TTGAAGCAAT ACTCATGCAA TTGTTTGT TTAAAGAAGA	7740
10	GGAAGAGGCA GAATCGATTG CATCACAAC AACAGATTGG ATAGAAACAT ATAAAAAGGA	7800
	GAAAGACTAA TGAACCTTAA GCAAGAAGTT GAGTCTAGAA AGACTTTTGC GATTATTTCA	7860
	CATCCCGATG CAGGGAAAAC AACGTTAACT GAAAACTAT TGTACTTCAG TGGTGCTATT	7920
15	CGTGAAGCGG GTACAGTTAA AGGGAAGAAG ACTGGTAAAT TTGCGACAAG TGAATGGATG	7980
	AAAGTTGAAC AAGAGCGTGG TATTTCTGTA ACTAGTTCAG TAATGCAATT TGATTACGAT	8040
20	GATTATAAAA TCAATATCTT AGATACACCA GGACATGAAG ACTTTTCAGA AGATACGTAT	8100
	AGAACATTAA TGGCAGTTGA CAGTGCTGTC ATGGTCATAG ACTGTGCAAA AGGTATTGAA	8160
	CCACAAACAT TGAAGTTATT TAAAGTTTGT AAAATGCGTG GTATTCCAAT CTTTACATT	8220
25	ATTAATAAAT TAGACCGAGT AGGTAAAGAA CCATTGAAT TATTAGATGA AATCGAAGAG	8280
	ACATTAAATA TTGAAACATA CCCTATGAAT TGGCCAATTG GTATGGGACA AAGTTTCTTT	8340
	GGCATCATTG ATAGAAAGTC TAAACAATT GAACCATTTA GAGATGAAGA AAATATATTA	8400
30	CATTTGAATG ATGATTTTGA GTTGAAGAA GATCATGCAA TTACAAATGA TAGTGATTTT	8460
	GAACAAGCGA TTGAAGAATT AATGTTGGTT GAAGAAGCGG GTGAAGCCTT TGATAATGAC	8520
	GCGCTGTTGA GTGGAGACTT AACACCTGTA TTTTTCGGTT CAGCTTTAGC TAACTTTGGT	8580
35	GTACAAAATT TCTTAAATGC ATATGTTGAT TTTGCGCCAA TGCCAAATGC GAGACAAACA	8640
	AAAGAGACG TTGAAGTAAG CCCGTTTGAT GATTCATTTT CAGGATTTAT CTTTAAATTT	8700
	CAAGCCAACA TGGACCCTAA ACACCGTGAT AGAATTGCCT TTATGCGTGT CGTTAGTGGT	8760
40	GCATTTGAAC GTGGTATGGA TGTACTTTG CAACGTACTA ATAAAAAGCA AAAGATCACA	8820
	CGTTCAACGT CATTTATGGC AGACGATAAA GAACTGTGA ATCATGCTGT AGCAGGCGAT	8880
	ATCATTGGAC TATATGATAC TGGTAATTAT CAAATTGGAG ATACTTTAGT TGGTGAAAA	8940
45	CAAACCTACA GTTTCCAAGA TTTACCACAA TTTACGCCAG AAATTTTAT GAAAGTTTCT	9000
	GCTAAAAACG TCATGAAACA GAAGCATTTT CATAAAGGTA TTGAACAATT AGTACAAGAA	9060
50	GGTGCGATTG AATACTATAA AACATTACAC ACAAAACCAA TTATTTTAGG TGCTGTTGGT	9120
	CAGTTACAAT TTGAAGTTTT CGAACATAGA ATGAAAAACG AATATAATGT TGATGTTGTT	9180

55

	AAGATGAACA CATCAAGATC GATTTTAGTG AAAGATAGAT ATGACGATTT AGTATTCTTA	9300
	TTTGAAATG AATTTGCAAC AAGATGGTTT GAAGAGAAAT TCCCTGAAAT TAAATTGTAT	9360
5	AGTTTACTTT AACAGCTCAA TTGTATAATC GAATTTGTTA CATTAAAAAT AATTGTTTCG	9420
	TTGAAGAAAA ATAAATTGTA TATTTTAAAA GAAAAAGGTA TACTATGATG TATCAAATGA	9480
10	ATAACCTATG GCATTTTGTC AGAGGGGAGT AACTTAAGAA TCATGACCGT ATAAATGaTT	9540
	CGACACTTTA TCGTCATTAC GArGATATCT TCCGGTAAAG TGGGCAATTT AAATTGCTTA	9600
	GTGAGACCTT TGCTATTTAT TTAGCATAGG TCTTTTTGTT TGTACTIONAAC TTATTTATTT	9660
15	AAAGGAGTTG TACATGTTAA TGGATCCAAG TTTGATCTTA CCTTATTTAT GGGTACTTGT	9720
	CGTTTTAGTA TTTTLAGAAG GCTTATTAGC AGCAGATAAC GCGATTGTTA TGGCTGTAAT	9780
	GGTTAAGCAC TTACCACCCG AACAAACGTAA AAAAGCTTTG TTTTACGGTT TGTTAGGTGC	9840
20	ATTTGTATTT AGATTTTLAG CATTATTCTT AATTAGTATT ATCGCGAACT TTTGGTTTAT	9900
	TCAAGCTGCA GGAGCGGTTT ACTTAATTTA TATGTCAATC AAAAATCTGT GGCAGTTCTT	9960
	TAAACACCCA GAAATTGAAA GTCCTGAAGC TGGAGATGAT CATCATTATG ATGAATCTGG	10020
25	TGAAGAGATT AAAGCAAGTA ACAAATCATT CTGGGGAAC TGTGTGAAAA TAGAATTTGC	10080
	AGATATCGCA TTTGCCATTG ATTCTATGCT TGCTGCTTTA gCTATTGCTG TAACACTTCC	10140
30	TAAAGTTGGT ATTCACTTTG GTGGTATGGA CTTAGGTCAG TTCGTAGTCA TGTTCCTAGG	10200
	TGGAATGATT GGTGTTATTC TAATGCGTTA TGCAGCAACA TGGTTTGTAG AGCTATTAAA	10260
	CAATATCCA GGACTTGAAG GTGCAGCCTt CGCGATCGTT GGTGGGTAG GTGTTAAATT	10320
35	AGTTGTCATG GTATTAGCGC ACCCAGACAT CGCTGTATTG CCTGAGCACT TCCCACATGG	10380
	CGTATTATGG CAATCTATTT TCTGGACAGT ACTAATTGGA TTAGTAATTA TCGGTTGGTT	10440
	AGGTTCAGTT GTTAAAAATA AAAAATCGCA TAAATAATTG ATGTGAAGCG GACAATCTTA	10500
40	ATTTAGTTTA AGGTTGTCCT TTTTCATTTA ATTGAGTGAT TTATGAAAAA TGGATTTTGA	10560
	AGAATGTGAA TCAAAAGATG CGATATAGTA TTAAGAAAAT GTGCCTTTTA TATTTAGCAT	10620
	TTTTTCAATA GAAATTATAT AGATTTTAAA GCAAATTAGG TGTTAATGTG TCATAATGAT	10680
45	AAGTGATTTT ATTGAATGGA GTGGACATTA GTGGATATTG GTAAAAACA TGTAATTCCT	10740
	AAAAGTCAGT nACCsaCGTA AGCGTCGTGA ATTCTTCCAC AACGAAGACA GAGAAGAAAA	10800
50	TTTAAATCAA CATCAAGATA AACAAAATAT AGATAATACA ACATCAAAAA AAGCAGATAA	10860
	GCAATACAT AAAGATTCAA TTGATAAGCA CGAACGTTTT AAAAATAGTT TATCATCGCA	10920
	TTTAGAACAG AGAAACCGTG ATGTTAATGA GAATAAGCT GAAGAAAGTA AAAGTAATCA	10980

55

	AAATTCATTA GATTCAGTGG ACCAAGATAC AGAGAAATCA AAATATTATG AGCAAAATTC	11100
	TGAAGCGACT TTATCAACTA AATCAACCGA TAAAGTAGAA TCAACTGAAA TGAGAAAGCT	11160
5	AAGTTCAGAT AAAAACAAAG TTGGTCATGA AGAGCAACAT GTACTTTCTA AACCTTCAGA	11220
	ACATGATAAA GAGACTAGAA TTGATTCTGA GTCTTCAAGA ACTGATTCAG ACAGCTCGAT	11280
	GCAGACAGAG AAAATAAAAA AAGACAGTTC AGATGGAAAT AAAAGTAGTA ATCTGAAATC	11340
10	TGAAGTAATA TCAGACAAAT CAAATACAGT ACCAAAATTG TCGGAATCTG ATGATGAAGT	11400
	AAATAATCAG AAGCCATTAA CTTTACCGGA AGAACAGAAA TTGAAAAGAC AGCAAAGTCA	11460
15	AAATGAGCAA ACAAAAACCT ATACATATGG TGATAGCGAA CAAAATGACA AGTCTAATCA	11520
	TGAAAATGAT TTAAGTCATC ATATACCATC GATAAGTGAT GATAAAGATA ACGTCATGAG	11580
	AGAAAATCAT ATTGTTGACG ATAATCCTGA TAATGATATC AATACACCAT CATTATCAAA	11640
20	AACAGATGAC GATCGAAAAAC TTGATGAAAA AATTCATGTT GAAGATAAAC ATAAACAAAA	11700
	TGCAGACTCG TCTGAAACGG TGGGATATCA AAGTCAGTCA ACTGCATCTC ATCGTAGCAC	11760
	TGAAAAAAGA AATATTTCTA TTAATGACCA TGATAAATTA AACGGTCAAA AAACAAATAC	11820
25	AAAGACATCG GCAAATAATA ATCAAAAAAA GGCTACATCA AAATTGAACA AAGGGCGCGC	11880
	TACGAATAAT AATTATAGTG ACATTTTGAA AAAGTTTGG ATGATGTATT GGCCTAAATT	11940
	AGTTATTCTA ATGGGTATTA TTATTCTAAT TGTTATTTTG AATGCCATTT TTAATAATGT	12000
30	GAACAAAAAT GATCGCATGA ATGATAATAA TGATGCAGAT GCTCaAAAAT ATACGACAAC	12060
	GATGAAAAAT GCCAATAACA CAGTTAAATC GGTCGTTACA GTTGAAAATG AAACATCAAA	12120
35	AGATTCTmTCA TTACCTAAAG ATAAAGCATC TCaAGACGAA GTGGGATCAG GTGTTGTATA	12180
	TAAAAAATCT GGAGATACGT TATATATTGT TACGAATGCA CACGTTGTCG GTGATAAAGA	12240
	AAATCaAAAA ATAACTTTCT CGAATAATAA AAGTGTTGTT GGGAAAGTGC TTGGTAAAGA	12300
40	TAAATGGTCA GATTTAGCTG TTGTTAAAGC AACTTCTTCA GACAGTTCAG TGAAAGAGAT	12360
	AGCTATTGGA GATTCAAATA ATTTAGTGTT AGGAGAGCCA ATATTAGTCG TAGGTAATCC	12420
	ACTTGGTGTA GACTTTAAAG GCACTGTGAC AGAAGGTATT ATTTCAGGTC TGAACAGAAA	12480
45	TGTTCTTATT GATTTGATA AAGATAATAA ATATGATATG TTGATGAAAG CTTTCCAAAT	12540
	TGATGCATCA GTAAATCCAG GTAACCGGG TGTGCTGTC GTCAATAGAG AAGGAAAATT	12600
50	AATAGGTGTA GTTGCAGCTA AAATTAGTAT GCCAAACGTT GAAAnTATGT CATTTGCA	12658

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6048 base pairs

55

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

	TGAAATnGAA TAGTACTATT GCAAGTGTAAGAGGTTAAT TTTTGCCnCA CGCGGGACTT	60
10	AAAAAGGCAA CCACTGGTTG TGACATATCC TTATTTACAT TTATAAATAT AAGGAGGAGG	120
	TAGTAGTGAA AGACTTATTG CAAGCACAGC AAAAGCTTAT ACCGGATCTC ATAGATAAAA	180
	TGTATAAACG TTTTCTATT CTTACTACTA TCTCAAAAA TCAGCCTGTC GGACGTCGAA	240
15	GTTTAAGCGA ACATATGGAT ATGACTGAAC GTGTACTGCG TTCTGAAACA GATATGCTTA	300
	AGAAACAAGA TTTGATAAAA GTTAAGCCTA CCGGAATGGA AATTACAGCT GAAGGTGAGC	360
	AACTGATTTT GCAATTGAAA GGTTACTTTG ATATCTATGC AGATGATAAT CGTCTGTCAG	420
20	AAGGTATTAA GAATAAATTT CAAATTAAGG AAGTTCATGT TGTTCTGGT GATGCTGATA	480
	ATAGTCAATC TGTTAAAACA GAATTAGGTA GACAAGCAGG TCAATTACTT GAAGGCATAT	540
	TACAAGAAGA CGCGATAGTT GCTGTAAGT GCGGATCCAC GATGGCATGT GTTAGTGAAG	600
25	CAATTCATTT ATTACCATAT AATGTATTCT TCGTACCAGC CAGAGGTGGA CTAGGCGAAA	660
	ATGTTGTCTT TCAGGCAAAC ACAATTGCAG CCAGTATGGC aCAACAAGCT GGCGGTTATT	720
30	ATACGACGAT GTATGTACCT GATAATGTCA GTGAAaCAAC ATATAATACA TTGTTGTTAG	780
	AGCCATCAGT CATAAACACT TTAGACAAAA TTAACAAGC AAACGTTATA TTACACGGCA	840
	TTGGTGATGC GCTGAAGATG GCGCATCGAC GTCAATCACC TGAAAAGGTC ATTGAACAAC	900
35	TTCAACATCA TCAAGCTGTC GGAGAGGCAT TTGGTTATTA TTTTGATACA CAAGGTCAAA	960
	TTGTCCATAA GGTTAAAACA ATTGGACTTC AATTAGAAGA CCTTGAATCA AAAGACTTTA	1020
	TTTTTGAGT TGCAGGAGGC AAATCGAAAG GTGAAGCAAT TAAAGCATAC TTGACGATTG	1080
40	CACCCAAGAA TACAGTGTTA ATCACTGATG AAGCCGCAGC AAAGATAATA CTTGAATAAG	1140
	AGATAAAAAG TTTAATACTT TTTAAATATC ATTTTAAAGG AGGCCATTAT AATGGCAGTA	1200
45	AAAGTAGCAA TTAATGGTTT TGGTAGAATT GGTCGTTTAG CATTCAGAAG AATTCAAGAA	1260
	GTAGAAGGTC TTGAAGTTGT AGCAGTAAAC GACTTAACAG ATGACGACAT GTTAGCGCAT	1320
	TTATTAAAT ATGACACTAT GCAAGGTCGT TTCACAGGTG AAGTAGAGGT AGTTGATGGT	1380
50	GGTTTCCGCG TAAATGGTAA AGAAGTTAA TCATTCAGTG AACCAGATGC AAGCAAATTA	1440
	CCTTGAAAG ACTTAAATAT CGATGTAGTA TTAGAATGTA CTGGTTTCTA CACTGATAAA	1500
	GATAAAGCAC AAGCTCATAT TGAAGCAGGC GCTAAAAAAG TATTAATCTC AGCACCAGCT	1560

55

	ACAGTTGTTT CAGGTGCTTC ATGTACTACA AACTCATTAG CACCAGTTGC TAAAGTTTTA	1680
	AACGATGACT TTGGTTTAGT TGAAGGTTTA ATGACTACAA TTCACGCTTA CACAGGTGAT	1740
5	CAAAATACAC AAGACGCACC TCACAGAAAA GGTGACAAAC GTCGTGCTCG TGCAGCGGCA	1800
	GAAAACATCA TCCCTAACTC AACAGGTGCT GCTAAAGCTA TCGGTAAAGT TATTCTTGAA	1860
10	ATCGATGGTA AATTAGATGG TGGTGACAA CGTGTTCCTG TAGCTACAGG TTCATTAACT	1920
	GAATTAACAG TAGTATTAGA AAAACAAGAC GTAACAGTTG AACAAAGTTAA CGAAGCTATG	1980
	AAAAATGCTT CAAACGAATC ATTCCGTTAC ACTGAAGACG AAATCGTTTC TTCAGACGTT	2040
15	GTAGGTATGA CTTACGGTTC ATTATTGAC GCTACACAAA CTCGTGTAAT GTCAGTTGGC	2100
	GACCGTCAAT TAGTTAAAGT TGCAGCTTGG TATGATAACG AAATGTCATA TACTGCACAA	2160
	TTAGTTCGTA CATTAGCATA CTTAGCTGAA CTTTCTAAAT AATTTTAGTA TAGTTTTTAT	2220
20	TCAAAATACGC TAGTGCTCAG AACTATTTAG CATTAAATTAA AGCTTATGAG TAAGCGGGGA	2280
	GCACAAACGC TTCTCCGCTT ATTTTATAT AAAATTTCTT AATTACAAGG AGGAAACACC	2340
	ATGGCTAAAA AAATTGTTTC TGATTTAGAT CTTAAAGGTA AACAGTCCT AGTACGTGCT	2400
25	GATTTTAACG TACCTTTAAA AGACGGTGAA ATTACTAATG ACAACCGTAT CGTTCAAGCT	2460
	TTACCTACAA TTCAATACAT CATCGAACAA GGTGGTAAAA TCGTACTATT TTCACATTTA	2520
30	GGTAAAGTGA AAGAAGAAAG TGATAAAGCA AAATTAACCT TACGTCCAGT TGCTGAAGAC	2580
	TTATCTAAGA AATTAGATAA AGAAGTTGTT TTCGTACCAG AAACACGCGG CGAAAAACTT	2640
	GAAGCTGCTA TTAAAGACCT TAAAGAAGGC GACGTATTAT TAGTTGAAAA TACACGTTAT	2700
35	GAAGATTTAG ACGGTAAAAA AGAATCTAAA AATGATCCAG AATTAGGTAA ATACTGGGCA	2760
	TCTTTAGGTG ATGTGTTTGT AAATGATGCT TTTGGTACTG CGCATCGTGA GCATGCATCT	2820
	AATGTTGGTA TTTCTACACA TTTAGAACT GCAGCTGGAT TCTTAATGGA TAAAGAAATT	2880
40	AAGTTTATTG GCGGCGTAGT TAACGATCCA CATAAACCAG TTGTTGCTAT TTTAGGTGGA	2940
	GCAAAAGTAT CTGACAAAAT TAATGTCATC AAAAACTTAG TTAACATAGC TGATAAAATT	3000
	ATCATCGGCG GAGGTATGGC TTATACTTTC TTAAGCGC AAGGTAAAGA AATTGGTATT	3060
45	TCATTATTAG AAGAAGATAA AATCGACTTC GCAAAAGATT TATTAGAAAA ACATGGTGAT	3120
	AAAATTGTAT TACCAGTAGA CACTAAAGTT GCTAAAGAAT TTTCTAATGA TGCCAAAATC	3180
50	ACTGTAGTAC CATCTGATTC AATTCCAGCA GACCAAGAAG GTATGGATAT TGGACCAAAC	3240
	ACTGTAAAAT TATTTGCAGA TGAATTAGAA GGTGCGCACA CTGTTGTATG GAATGGACCT	3300
	ATGGGTGTAT TCGAGTTCAG TAACTTTGCA CAAGGTACAA TTGGTGTATG TAAAGCAATT	3360

55

	TCTTTAGGTT TTGAAAATGA CTTCACTCAT ATTTCAACTG GTGGCGGCGC GTCATTAGAG	3480
	TACCTAGAAG GTAAAGAATT GCCTGGTATC AAAGCAATCA ATAATAAATA ATAAAGTGAT	3540
5	AGTTTAAAGT GATGTGGCAT GTTTGTTTAA CATTGTTACG GGAAAACAGT CACAAGATGA	3600
	CATCGTGTTC CATCACTTTT CAAAAATATT TACAAAACAA GGAGTGTCTT TAATGAGAAC	3660
10	ACCAATTATA GCTGGTAACT GGAAAATGAA CAAAACAGTA CAAGAAGCAA AAGATTCGTC	3720
	AATACATTAC CAACACTACC AGATTCAAAA GAAGTAGAAT CAGTAATTG TGCACCAGCA	3780
	ATTCAATTAG ATGCATTAAC TACTGCAGTT AAAGAAGGAA AAGCACAAGG TTTAGAAATC	3840
15	GGTGCTCAAA ATACGTATTT CGAAGATAAT GGTGCGTTCA CAGGTGAAAC GTCTCCAGTT	3900
	GCATTAGCAG ATTTAGGCGT TAAATACGTT GTTATCGGTC ATTCTGAACG TCGTGAATTA	3960
	TTCCACGAAA CAGATGAAGA AATTAACAAA AAAGCGCACG CTATTTTCAA ACATGGAATG	4020
20	ACTCCAATTA TATGTGTTGG TGAAACAGAC GAAGAGCGTG AAAGTGGTAA AGCTAACGAT	4080
	GTTGTAGGTG AGCAAGTTAA GAAAGCTGTT GCAGGTTTAT CTGAAGATCA ACTTAAATCA	4140
	GTTGTAATTG CTTATGAACC AATCTGGGCA ATCGGAACTG GTAAATCATC AACATCTGAA	4200
25	GATGCAAATG AAATGTGTGC ATTTGTACGT CAAACTATTG CTGACTTATC AAGCAAAGAA	4260
	GTATCAGAAG CAACTCGTAT TCAATATGGT GGTAGTGTTA AACCTAACAA CATTAAAGAA	4320
30	TACATGGCAC AAAGTATAT TGATGGGGCA TTAGTAGGTG GCGCATCACT TAAAGTTGAA	4380
	GATTCGTAC AATTGTTAGA AGGTGCAAAA TAATCATGGC TAAGAAACC AACTGCGTTAA	4440
	TTATTTTAGA TGTTTTTGGC AACCGCGAAA GCGAACATGG TAATGCGGTA AAATTAGCAA	4500
35	ACAAGCCTAA TTTTGATCGT TATTACAACA AATATCCAAC GACTCAAATC GAAGCGAGTG	4560
	GCTTAGATGT TGGACTACCT GAAGGACAAA TGGGTAACTC AGAAGTTGGT CATATGAATA	4620
	TCGCTGCAGG ACGTATCGTT TATCAAAGTT TAACTCGAAT CAATAAATCA ATTGAAGACG	4680
40	GTGATTTCTT TGAAAATGAT GTTTTAAATA ATGCAATTGC ACACGTGAAT TCACATGATT	4740
	CAGCGTTACA CATCTTTGGT TTATTGTCTG ACGGTGGTGT ACACAGTCAT TACAAACATT	4800
	TATTTGCTTT GTTAGAACTT GCTAAAAAAC AAGGTGTTGA AAAAGTTTAC GTACACGCAT	4860
45	TTTAGATGG CCGTGACGTA GATCAAAAAT CCGCTTTGAA ATACATCGAA GAGACTGAAG	4920
	CTAAATTCAA TGAATTAGGC ATTGGTCAAT TTGCATCTGT GTCTGGTCGT TATTATGCAA	4980
50	TGGATCGTGA CAAACGTTGG GAACGTGAAG AAAAAGCTTA CAATGCTATT CGTAATTTTG	5040
	ATGCCCCAAC TTATGCAACT GCCAAGAAG GTGTAGAAGC AAGCTATAAT GAGGGCTTAA	5100
	CTGACGAATT CGTAGTACCA TTCATCGTTG AGAATCAAAA TGACGGTGTT AATGATGGAG	5160

55

5 CGAACAGAGC ATTCGAAGGC TTTAAAGTTG AACAAAGTTAA AGACTTATTC TATGCAACAT 5280
 TCACTAAGTA TAATGACAAT ATCGATGCGG CTATCGTCTT CGAAAAAGTT GATTTAAATA 5340
 10 ATACAATTGG TGAAATTGCA CAAAATAACA ATTTAACTCA ATTACGTATT GCAGAAACTG 5400
 AAAAATACCC TCACGTTACT TACTTTATGA GTGGTGGACG TAACGAGGAA TTTAAAGGTG 5460
 AACGCCGTCG TTAAATTGAT TCACCTAAAG TTGCAACGTA TGACTTGAAA CCAGAAATGA 5520
 GTGCTTATGA AGTTAAAGAT GCATTATTAG AAGAGTTAAA TAAAGGTGAC TTGGACTTAA 5580
 TTATTTTAAA CTTTGCTAAC CCTGATATGG TTGGACATAG TGGTATGCTT GAGCCGACAA 5640
 15 TCAAAGCAAT CGAAGCGGTT GATGAATGTT TAGGAGAAGT GGTTGATAAG ATTTTAGACA 5700
 TGGACGGTTA TGCAATTATT ACTGCTGACC ATGGTAACTC TGATCAAGTA TTGACGGaTG 5760
 ATGATCAACC AATGACTACG CAwACAACGA ACCCAGTACC AGTGATTGTA ACAAAGAAG 5820
 20 GCGTTACACT TAGAGAAACT GGTGCTTAG GTGACTTAGC ACCTACATTA TTAGATTTAT 5880
 TAAATGTAGA ACAACCTGAA GATATGACAG GTGAaTCTTT AATTAAACAC TAATATTGTA 5940
 AAAGATGTTA AGTAAACGCT TAATGACACT TATTTTTTGA AAATAATAGT AATATCnTTT 6000
 25 TGTAAATGA AAGAATAAAG CTATAATAAT TATAGAATAA CTATTTan 6048

(2) INFORMATION FOR SEQ ID NO: 129:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5602 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

AAAGAAAGTGC AAGATATCAT CGCATTAAAT AAGTCGTTAC AAAGTGTAAT TGTAGACaTC 60
 40 GCTTCCAATA ATGTTGATAC AATTATGCCT GGTTATACTC ATTTACAGCG TGCACAGCCA 120
 ATTTCAATTG CACATCATAT TATGACTTAT TTTTGGATGT TACAACGAGA CCAACAACGA 180
 TTTGAAGATA GTTTAAAACG AATCGATATT AATCCTTTAG GTGCAGCAGC CTTAAGTGGT 240
 45 ACCACATACC CTATCGATAG ACACGAGACA ACAGCATTGT TGAACTTTGG CAGTCTCTAT 300
 GAGAATAGCC TAGATGCTGT TAGTGACAGA GACTATATTA TTGAAACATT GCATAATATT 360
 TCTTTAACGA TGGTTCACCT ATCACGCTTT GCAGAGGAAA TTATTTTCTG GTCCACAGAC 420
 50 GAAGCTAAAT TCATTACATT ATCAGATGCA TTTTCAACTG GCTCATCTAT TATGCCACAA 480
 AAGAAAAATC CTGATATGGC AGAATTAATT AGAGGTAAAG TTGGTCGAAC GACTGGTCAT 540

55

	GAAGATAAAG AAGGTTTATT CGATGCTGTC CATACAATTA AAGGTTCTTT ACGTATTTTC	660
	GAAGGTATGA TTCAAACGAT GACAATTAAT AAAGAACGAC TCAATCAAAC TGTAAAGAA	720
5	GATTTTTCATG ATGCAACGGA ACTAGCAGAT TATTTAGTAA CTAAAAATAT TCCATTTAGA	780
	ACTGCACATG AAATTGTAGG AAAAATCGTC TTAGAATGTA TACAACAAGG TCATTATTTA	840
10	TTAGATGTTT CTTTAGCAAC ATATCAACAA CATCATTTCTA GTATTGATGC CGATATTTAC	900
	GATTATTTGC AGCCTGAAAA TTGTTTAAAA CGACGTCAA GTTACGGTTC AACAGGTCAA	960
	TCATCGGTCA AACAACTT TGATGTTGCT AAACAATTAC TATCACAATA AATACGTTAA	1020
15	TCTACCTACC CACAATGTCT ATTAAAATTA CATTTGTTGGT ATTTTAATGC TCTCTTCGTC	1080
	TTGTTGAACA TCACATTTT AAGATTCCTA AAATGTTTGA TAATTCCTTT AAATTTATAT	1140
	TACAAAAATG TTATAAATTG TAAAAGAAAT GTGTAAAGCG TTTTCACAAG CAGGTTTTTG	1200
20	TAGTATTTTA AAATTGTTAG ACTACAAATA AAGAGATGAA AGGATAAAGA CTATGACTAA	1260
	CTCTTCGAAA AGCTTCACTA AATTTATGGC TGCTTCTGCT GTTTTACTA TGGGATTTT	1320
	ATCAGTACCT ACTGCTGGCG CTGAACAAAC AAATCAAATT GCAAATAAAC CTCAGGCTAT	1380
25	TCAATGGCAT ACAAATTTAA CGAATGAGCG ATTCACTACT ATCGCACATC GTGGCGCAAG	1440
	TGGCTATGCA CCCGAGCATA CGTTTCAAGC ATATGATAAG AGTCATAATG AGTTAAAAGC	1500
30	ATCTTATATC GAAATTGATT TACAACGTAC CAAAGATGGC CATTTAGTTG CTATGCATGA	1560
	TGAAACTGTT AACCGTACAA CAAATGGACA CGGTAAAGTT GAGGATTATA CCCTTGATGA	1620
	ATTAAAACAG TTAGATGCAG GAAGTTGGTT TAATAAAAAA TATCCAAAAT ACGCAAGAGC	1680
35	AAGTTATAAA AATGCTAAAG TACCCACTTT AGATGAAATT TTAGAACGTT ATGGCCCGAA	1740
	TGCAAACTAT TATATTGAAA CAAAGTCACC TGATGTATAC CCAGGAATGG AAGAACAATT	1800
	ATTAGCTTCA TTGAAAAAGC ATCACCTTTT AAATAACAAT AAATTAAAAA ATGGACATGT	1860
40	AATGATTCAA TCATTTTCTG ACGAAAGTTT AAAGAAAATT CATCGTCAA ATAAGCATGT	1920
	GCCATTAGTA AAATTAGTTG ATAAAGGTGA ACTACAACAA TTTAACGACC AACGCTTAAA	1980
	AGAGATACGC TCTTATGCGA TTGGATTAGG TCCTGATTAT ACAGATTTAA CTGAACAAAA	2040
45	TACCCATCAT TTAAAAGACT TAGGATTTAT AGTACATCCT TATACAGTGA ATGAAAAAGC	2100
	TGATATGTTA CGATTAAATA AATATGGCGT TGATGGTGTC TTTACAAATT TCGCTGATAA	2160
50	ATATAAAGAA GTCATTAAGT AGTAATGTTA AACTAGAAAA CATAAATACA AAAATATAGC	2220
	TATTACTATA AAAACAGCA GTAAGATATT TCCAAATTGA AATTATCCTA CTGCTGTCTT	2280
	TTTGGGAGTG GGACAGAAAT GATATTTTCG CAAAATTTAT TTCGTCGTCC CACCCCAACT	2340

55

	TTGTCTGTAG AAATTGAGGA GCTAATTTCT CTGTGTCGGG GCTCCACCCC AACTTGACACA	2460
	CTATTGTAAG CTGACTTTCC GCCAGCCTCT GTGTTGGGGC CCCGCCAACT TGCACACTAT	2520
5	TGTAAGCTGA CTTTCCACCA GCCTCTGTGT TGGGGCCCCG ACTATTTTGT AAAAGAGCGT	2580
	GTTACACGGG CATTGTTTTA CAGTCAACTA CTGCTAAAAT AAAATTAACG AGCTTAGGGC	2640
10	TTTGTCTTCT GTCCCAAGCT CGTTAAATCA CATATGATAA TTAATTATGC CCAACCACGA	2700
	TATCTAGCTG CTTCTGCTGT ACGTTTAATA CCTATGATAT ATGCTGCAAG TCTCATATCT	2760
	ATTTTTCGGT TTTGAGACAA TTCGTAAATC GTATCAAATG CCGCTTCTAA TTTTTCACGT	2820
15	AGCTTTTCAT TAACCTCTTC TTCAGACCAA TAATAACCTT GATTATTTTG TACCCATTCTG	2880
	AAGTAAGAAA CCGTACACC ACCAGCACTT GCTAATACGT CTGGAACATA TAATATACCA	2940
	CGTTCAGTTA AAATACGTGT TGCTTCTGGT GTTGTAGGTC CATTAGCAGC TTCAACAACG	3000
20	ATACTAGCTT TAATATCATG TGCATTGTCT TCTGTAATTT GGTTTGAAAT AGCCGCTGGT	3060
	ACTAAAATGT CACAATCTAA TTCAAACAAT TCTTTATTTG AGATTGTTTC TTCAAATAAA	3120
	TTTGTTACCG TACCAAAAC ATCACGACGG TCTAATAAAT AATCTATATC TAAGCCATTT	3180
25	GGATCGTGTA ATGCACCGTA AGCATCAGAG ATACCTACAA TTTTTCACC TAAATCATAT	3240
	AAGAATTTAG CTAAGAAACT TCCGGCATT ACGAAACCTT GAATAACAAC CTTGGCACCT	3300
30	TCAATTTGCA TATTACGACG TTTTGCAGCT TGTTC AATTG CAATAACTAC ACCTAGTGCA	3360
	GTTGATCTGT CGCGTCCATG AGAACCACCC AATACAATTG GTTTACCTGT GATGAAACCT	3420
	GGTGAATTAA ATTTATCTAA TGCACTATAT TCATCCATCA TCCAAGCCAT AATTTGTGAG	3480
35	TTTGTAATA CATCTGGTGC TGGAAATATCT TTGTTCCGAC CTACGAATTG TGAAATTGCT	3540
	CTTACATATC CGCGTGATAA ACGTTCAACT TCATGAATGC TCATTTGACG TGGATCACAA	3600
	ACGATACCAC CCTTACCACC ACCGTATGGT AAGTTTACAA TGCCACATTT CAAAGTCATC	3660
40	CACATTGATA ATGCTTTTAC TTCTTCTTCA TCAACATCTG GGTGGAAACG CACGCCCCCT	3720
	TTTGTTGGTC CAACAGCATC ATTATGTTGC GCACGGTAAC CTGTGAATGT TTTTACTGTG	3780
	CCATCATCCA TTCGTACAGG GATACGCACT TGTAACATTC TTAAAGGTTT TTTAATTAAA	3840
45	TCGTACATTC CTTCGTCAAA TCCCAATTTA TGCAATGCTT CTTTAATAAT TCCTTGAGTA	3900
	GAAGTTACTA AATTATTGTT CTCAGTCATG ATCCTTTTCG CCTCTTCTTT ACCTAATGAT	3960
50	TTCGCTTTCA AACATATTGT AACATAACGT ATTCTTTTTT AAAGCCCTTA CAACTGATT	4020
	GTTACAACCTT TTTGACATTA TTGAAATACA TGTCTTATTT TTTCAAGTGC AAGGTCCAAT	4080
	TCTTCTTTAG TAATAATTAA TGGTGGTGCA AAACGAATGA CAGTATCATG CGTTTCTTTA	4140

55

ACACCTATAA ACAAACCACG TCCACGGACT TCTTTAATTG ATGGATGATC AATTTGCTTT 4260
 AATTGTTCTT TAAAATAATC TCCTAATTCT AAAGAGCGGC CTGGTAAATC CTCATCAACG 4320
 5 ATAACATCTA ATGCAGCAAT TGATGCAGCA CAAGCAAGTG GATTACCACC AAATGTTGAA 4380
 CCATGTGAGC CAGGTGTAAA GACATCTAAT ACTTCTTTAT CTGCTAATAC AACAGAAATT 4440
 GGAAGACTC CACCACCTAG TGCTTTACCT AAAATATAGA CATCAGGTTT TACATTATCC 4500
 10 CAATCCGTAG CAAATAATTT ACCCGAACGA CCTAATCCTG CTTGGATTTT GTCAGCAATA 4560
 AATAAGACAT TATGTTTCATC ACATAATTCT CTAATTGCTT TCAAATATCC TTCTGGCGGT 4620
 ATATTTATAC CCGCTTCACC TTGAATTGGT TCTACTAAAA CTGCTGCAGT ATTTTCATTA 4680
 15 ATTGCAGCTT TCAATGCATC TACATCTCCA AAATCAACTT TTCTAAATCC ATCTAATAAC 4740
 GGACCATAAC CACGTTGGTA TTCTGCTTCT GAAGATAATG AAAGTGGCGC CATTGTTGCA 4800
 20 CCATGGAAGT TACCATTAAA TGCAATGATT TCTGCTTTAT TTGGCTCAAT TCCTTTAACA 4860
 TCGTATGCC AGCGTCGTGC TGCTTTCAA GCTGTTTCTA CTGCTTCAGC ACCTGTATTC 4920
 ATTGGTAAAG CTTTATCTTT ACCTGCCAGT TTACAAATTT TTTCGTACCA TTCACCTAAG 4980
 25 TTATCACTAT GAAAAGCAGC TGAACTAAA GTCACCTTAT CAGCTTGATC TTTTAATGCT 5040
 TGAATAATTT TCGGATGTCT ATGACCTTGG TTAACAGCGG AATATGCAGA TAACATATCC 5100
 ATATATTTAT TGCCTTCAGG ATCTTTAACC CATACCCCTT CAGCTTcTGa AATGaCAATT 5160
 30 GGcAATGGTA AATAATTATG TGCTCCGTAA TGATTGTGTA ACTCAATAAT TTTTTCAGAT 5220
 TTAGTCATCA TATCTCCCTT TTTTCATCATT TATAACTATT ATACATGAAA CATTATCCAA 5280
 ATAATTACAT TAGTTTTCAA AGCAGATACT TTTCCACCAA AAAAGATGAA ATAATCACTA 5340
 35 AGTTTCATTA AATTTGTCTA TTTTGAAAAC CCTTACATTT ATAATGACAT AATTACTTAA 5400
 ATGaJTACAA GCAAAGAAT TGATAATTTT ACACCTAATC AAAAGTATAT TTTACTAAGA 5460
 ATATTTTTAT TTATAAATAT TGAAAACCAC TAACAAATTG CATAACAAT ATCATTAGTG 5520
 40 GTAACAGTTA AACACTTATT TATCTTTACG GGGTAATGGG TTAAAACCCT TnCATTAAAA 5580
 TTGGATGnCC ATAAAATTAG GG 5602

45 (2) INFORMATION FOR SEQ ID NO: 130:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5924 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

55

	TAACCCCATTTTACCTGGAA AAATCgTTTG CGATGCaATm GCaTtGaAT ATAAATACAT	60
	TTTACGTATa GAATTATAAA AgGTTTCATT CaAATCTTAG GGTCAAAAAT GTTATAATAT	120
5	TTTTATGTCA AATTTAAAAC AGTAACACTT ATTTACAAGG TTGCAATATT TTGAAGTAAT	180
	AAAGGAAGTG TCGCGTATTT TAACTTTTTTC AGAGCAAAAT GCACTCGCGA AAATAGATGA	240
10	TTTAATGAAT ACTTATTGCA ATCAATGTCC AATCAAACT CGTCTGCGTA AATTAGAGGG	300
	GAAAACGAAG GCGCATCATT TTTGTATCAA TGAGTGTTCa ATAGGGAAAG AAATAAAACA	360
	ATTAGGAAAT GAACTTCAAT AGGAGGAAGT CAAATGAAAA TTATATCTAT ATCAGAAACA	420
15	CCGAACCACA ACACAATGAA GATTACACTT AGTGAAAGCA GAGAAGGTAT GACATCAGAT	480
	ACGTATACTA AAGTTGATGA TTCACAGCCA GCATTTATTA ATGACATCTT AAAGGTTGAA	540
	GGCGTTAAAT CAATTTTCCA TGTTATGGAC TTTATTTTCAG TAGATAAAGA AAATGACGCA	600
20	AATTGGGAAA CAGTATTGCC AAAAGTAGAG GCTGTATTTCG AATAAATTTT TCATCAACTA	660
	GTATTCGGGG GGAATAAAGT ATATGGAAAT TTTACGTATA GAGCCAACAC CAAGTCCAAA	720
	TACAATGAAA GTTGTTTTGT CATATACAAG AGAAGACAAG TTATCTAATA CTTATAAAAA	780
25	AGTAGAAGAA ACACAACCAA GATTTATAAA TCAGTTGTTA TCTATAGATG GTATCACTTC	840
	CAATTTTCAT GTCATGAACT TCTTAGCTGT TGATAAGGCA CCAAAGCTG ATTGGGAAGT	900
	CATATTACCT GATATTAAAG CTGCTTTTTTC TGATGCGAAT AAGGTTTTAG AATCTGTAA	960
30	TGAACCTCAA ATTGACAATC ATTTTGGTGA AATTAAAGCT GAATTATTAA CTTTAAAGGG	1020
	TATACCGTAT CAAATTAAGC TAACTTCTGC TGACCAAGAA TTAAGAGAAC AATTACCACA	1080
35	AACATATGTT GACCATATGA CTCAAGCGCA AACAGCACAT GACAATATTG TTTTATGCG	1140
	TAAATGGCTA GATTTAGGAA ATCGCTATGG AAATATTCAA GAAGTAATGG ATGGTGTCTT	1200
	AGAAGAAAGTG CTAGCTACCT ATCCAGAATC ACAGTTACCC GTATTGGTAA AACATGCTTT	1260
40	AGAAGAAAAT CACGCAACTA ATAATTATCA TTTCTATCGA CATGTCTCTT TGGATGAATA	1320
	TCATGCAACT GATAATTGGA AGACTCGATT ACGAATGTTA AACCATTTTC CAAAGCCGAC	1380
	TTTTGAAGAT ATACCGCTGC TTGATTTAGC TTTATCTGAT GAAAAAGTAC CGGTTAGACG	1440
45	TCAAGCGATT GTATTATTAG GTATGATTGA AAGTAAAGAA ATTTTACCGT ATTTATATAA	1500
	GGGGCTTCGT GATAAAAGTC CTGCTGTAAG AAGAACAGCA GGGGATTGCA TAAGCGATTT	1560
	AGGGTATCCA GAGGCACTAC CAGAAATGGT GCTACTATTA GATGATCCAC AGAAAATCGT	1620
50	TAGGTGGCGT GCTGCTATGT TTATCTTTGA TGAAGGTAAT GCAGAGCAGC TTCCCGCACT	1680
	AAAAGCCCAT ATTAATGACA ATGCGTTTGA AGTTAAATTA CAAATTGAAA TGGCCATATC	1740

55

	AATTTAATTG GAGGAATTAA ATATGAATGC ATATGATGCT TATATGAAAG AAATTGCGCA	1860
	ACAAATGCGT GCGGAATTAA CTCAAAATGG TTTTACAAGT TTAGAAACGA GCGAACAGct	1920
5	ATCGGAGTAT ATGAACCAAG TAAATGCTGA TGACACTACT TTTGTAGTTA TTAACCTCTAC	1980
	ATGCGGCTGT GCAGCTGGAT TAGCAAGACC AGCTGCAGTA GCAGTTGCAA CACAAAATGA	2040
10	ACATAGACCT ACAAATACAG TTACAGTTTT TGCTGGGCAA GATAAAGAAG CAACTGCTAC	2100
	AATGCGAGAA TTCATTCAGC AAGCACCATC TAGTCCTTCG TATGCTTTAT TCAAAGGTCA	2160
	AGATTTAGTT TATTTTATGC CTAGAGAATT TATCGAAGGT AGAGATATTA ATGACATTGC	2220
15	AATGGACTTA AAGGATGCCT TTGACGAAAA TTGTAAATAG TACACATAAA TAAATATAAA	2280
	GGTTAACACA TTTTATAATA TTAAAAATGG TGTCTGTCAT TGAAAATAGA GAATATAGTT	2340
	GTATTCTATT TGTTAAATAA AGTCCGTTTT TACCaaCTAT ATTTTCTAGA AATTAACTG	2400
20	TTTTAATAGG ACATCAAACA TAATATTCaA ATCaTGTGTT AACCTCTTTT TTAAAATTTT	2460
	TTAGCATTAA AGTTATAGAT TTGGGTAAAC AATTACCAAT TGGAAACATA TATCACGTTA	2520
	CGATGGGGTA GGTACTTAAT CAGCATTTTA TAAATAAAGT AACGGAATTC ATGATATTAA	2580
25	TATCATATTC CTAAAATGAG TGATAACAAA ATGCTACATA AAGTTAAGTT ATATCAAAC	2640
	AAATATACAT ACTATAAATA ATGAAAATGA GGTGTTATCG CATATGTTGA ATTCATTTGA	2700
	TGCAGCATAT CACAGTCTTT GTGAAGAAGT TTTAGAAATA GGAAATACAC GAAATGATCG	2760
30	CACAAATACA GGTACGATTT CGAAATTTGG TCATCAACTT CGCTTTGACT TATCTAAAGG	2820
	ATTTCCACTA TTAACGACAA AGAAAGTTTC TTTTAAATTA GTAGCAACCG AATTATTATG	2880
35	GTTTCATTAA GGAGATACAA ACATCCAATA CTTATTAAAA TATAATAATA ATATATGGAA	2940
	CGAATGGGCT TTTGAAAATT ATATCAAATC AGACGAGTAT AAAGGTCCAG ATATGACAGA	3000
	TTTCGGGCAT CGTGCAATTGA GTGATCCTGA ATTAAACGAA CAATATAAAG AACAAATGAA	3060
40	ACAATTAAAG CAACGTATTC TTGAAGATGA TACATTTGCG AAGCAATTTC GGGATTTAGG	3120
	AAATGTTTAT GGTAAACAAT GGCGAGATTG GGTGATAAA GATGGTAATC ATTTTGATCA	3180
	ACTTAAACA GTAATTGAAC AAATTAAGCA TAATCCAGAT TCAAGGCGAC ACATCGTATC	3240
45	TGCATGGAAT CCAACAGAAA TTGATACAAT GGCACCTCCG CCTTGTCTATA CCATGTTCCA	3300
	GTTTTATGTC CAAGATGGTA AGTTAAGTTG CCAGTTATAC CAACGTAGCG CAGATATCTT	3360
	TTTAGGTGTG CCATTTAATA TCcGCagctA CGCTTTATTG ACACACCTTA TTGCCAAAGA	3420
50	ATGTGGACTT GAAGTGGGTG AATTTGTGCA TACATTTGGA GATGCACATA TTTATTCAAA	3480
	TCATATTGAT GCGATTCAAA CACAATTAGC ACGTGAAAGC TTCAATCCTC CAACATTAAA	3540

55

	TGAATCACAT CCAGCAATAA AAGCTCCAAT AGCAGTGTAG TCATTGCATA GTTAGCTAAC	3660
	CATATAGACA TCAAAATGAC ATCATAGTAT TTTCAAGTGC AAAAAAGTAC TTTTTTGTGT	3720
5	TAAACGTTTT CATAAATTAT GCAAAATCAT TATTCTATC ACACTTTATG ATAAAAATTG	3780
	TGTTAAATTA AAGATAACTT AGTAATAAAA AATGAAATGA TAGAAGAAGG AGGATAATTA	3840
	TGACTTTATC CATTCTAGTt GCACATGACT TGCAACGAGT AATTGGTTTTt GAAAATCAAT	3900
10	TACCTTGGcA CCTACCAAAT GATTTGAAGC ATGTTAAAAA ATTATCAACA GGTCATACTT	3960
	TAGTAATGGG TCGTAAGACA TTTGAATCGA TTGGTAAACC ACTACCGAAT CGTCGAAATG	4020
	TTGTACTTAC TTCAGATACA AGTTTCAACG TAGAnGGCGT TGATGTAATT CACTCTATTG	4080
15	AAGATATTTA CCAACTACCG GGCCATGTTT TCATATTTGG AGGGCAAACA TTATTTGAAG	4140
	AAATGATTGA TAAAGTGGAC GACATGTATA TTA CTGTTAT TGAAGGTAAA TTCCGTGGTG	4200
20	ATACGTTCTT TCCACCTTAT mCATTkGagr CTGGGAAGTT GCCTCTTCAG TTGAAGGTAA	4260
	ACTAGATGAG AAAAAATACAA TTCCACATAC CTTTCTACAT TTAATTCGTA AAAAAAAGG	4320
	GGGAAAACGA CCATGACAAA ACAGATTATA GTAACAGACT CAACATCCGA TTTATCTAAA	4380
25	GAATACTTAG AAGCAAACAA CATTCAATGTA ATTCTTTTAA GTTTAACTAT TGAAGGAGCT	4440
	TCATACGTTG ACCAAGTAGA TATTACATCA GAAGAATTTA TTAATCATAT TGAAAATGAT	4500
	GAAGATGTAA AGACAAGTCA GCCAGCCATA GGTGAATTTA TATCTGCTTA TGAAGAACTA	4560
30	GGAAAAGATG GCTCTGAAAT CATAAGTATT CATCTTTCTT CAGGATTAAG TGGTACATAT	4620
	AACACTGCTT ACCAAGCAAG TCAAATGGTA GATGCTAATG TAACTGTTAT TGATTCAAAA	4680
35	TCTATTTCTT TTGGTTTtagG GTATCAAATA CAACACCTAG TAGAGCTTGT AAAAgAaGGT	4740
	GtCTCAACTT CTGAAATAGT TAAAAAGTTA AATCATTTAA GAGAAAACAT TAAATTATTT	4800
	GTAGTTATAG GGCAATTGAA TCAATTAATT AAAGGTGGCA GAATTAGTAA AACAAAAGGT	4860
40	TTGATTGGTA ATCTTATGAA AATTAAACCA ATTGGTACAC TAGATGATGG TCGCTTAGAG	4920
	CTTGTGcmCA ATGCGAGAAC TCaAAATTck AGTATCCAAT ACTTGAAAAA GGAAATTGCT	4980
	GAATTTATAG GAGATCATGA AATCAAATCC ATTGGTGTcG CACATGCTAA CGTCATTGAA	5040
45	TATGTTGATA AATTGAAGAA AGTTTTTAAT GAAGCTTTTC ATGTGAATAA TTACGATATA	5100
	AATGTAAC TA CACCAGTTAT TTCTGCACAT ACTGGTCAAG GTGCGATTGG CcTCGTAGTC	5160
	CTTAAGAAGT AAATTTAATC TTTTCAGTGT TAATTACTTC CATTTCaATC CTTTATAGAC	5220
50	TAAATTTATA ATTAGATAGA TAGAGGAGGT AATTCATATG ACAAAGAAT ATGCAACATT	5280
	AGCAGGAGGA TGTTTTCTGGT GCATGGTTAA ACCATTTACA TCATATCCAG GCATCAAGTC	5340

55

GAATCAAACC GGCCATGTCG AAGCAGTACA AATTACGTTT GATCCAGAGG TTACTTCCTT 5460
 TGAAATATA TTAGACATAT ATTTCAAAC ATTTGACCCA ACTGATGATC AAGGGCAATT 5520
 5 TTTGATAGA GCGGAAAGCT ATCAACCAGT CATTTTCTAT CATGATGAAC ATCAGAAAAA 5580
 GGCTGCTGAG TTTAAAAAGC AACCAATTAA TGAACAAGGT ATTTTCAAGA AACCAGTGAT 5640
 TACACCTATT AAACCATATA AAAATTTCTA TCCAGCTGAA GACTACCATC AAGATTATTA 5700
 10 CAAAAAGAAC CCGGTACATT ATTACCAATA TCAACGTGGT TCAGGTAGAA AAGCGTTTAT 5760
 AGAATCACAT TGGGGGAATC AAAATGCTTA AAAAGATAA AAGTGAAC TAACAGATATAG 5820
 AATATATTGT TACACAAGAA AACCGCACTG AACCACCATT TATGAATGAA TATTGGAATC 5880
 15 ATTTTGCTAA AGGATTTATG TAGATAAAAT TCnGGTAAAC CTTG 5924

(2) INFORMATION FOR SEQ ID NO: 131:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9280 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

GGCCGTTnAA AATCTCCAAA ATAnAAAAAC CCATCTTGTT CCAATGTTTT AAAATCGCCa 60
 30 TCCaACACTT GaTCaATAGC TTGCAACAAC GTTGAACGTG TTTTaCCAAA AGCATCaAAC 120
 GCTCCCACTA AAATCAGTGC TTCAAGTAAC TTTCTCGTTT TGA CTCTCTT CGGTATACGT 180
 35 CTAGCAAAAT CAAAGAAATC TTTAAATTTG CCGTTCTGAT AACGTTTCATC AACAATCACT 240
 TTCACACTTT GATAACCAAC ACCTTTAATT GTACCAATTG ATAAATAAAT GCCTTCTTGG 300
 GAAGGTTTAT AAAACCAATG ACTTTCGTTA ATGTTCTGGTG GCAATATAGT GATACCTTGT 360
 40 TTTTTTGCTT CTCTATCAT TTGAGCAGTT TTCTTCTCAC TTCCAATAAC ATTACTTAAA 420
 ATATTTGCGT AAAATAAATT TGGATAATGG ACTTTTAAAA AGCTCATAAT GTATGCAATT 480
 TTAGAATAGC TGACAGCATG TGCTCTAGGA AAACCATAAT CAGCAAATTT CAGAATCAAA 540
 45 TCAAATATTT GCTTACTAAT GTCTTCGTGA TAACCATTTT GCTTGCACC TTCTATAAAA 600
 TGTGACGCT CACTTTCAAG AACAGCTCTA TTTTTTTTAC TCATTGCTCT TCTTAAATA 660
 TCCGCTTCAC CATAACTGAA GTTTGCAAT GTGCTCGCTA TTTGCATAAT TTGCTCTTGA 720
 50 TAAATAATAA CACCGTAAGT ATTTTTTAAT ATAGGTTCTA AATGCGGATG TAAATATTGA 780
 ACTTTGCTTG GATCATGTCT TCTTGTAATG TAAGTTGGAA TTTCTTCCAT TGGACCTGGT 840

	ACACTTCTTA CACCGTCAGA CTCTAATTGG AATATGCCAG TCGTATCTCC TTGCGACAAC	960
	AATTCAAACA CTTTTTGATC ATCAAACGGA ATCTTTTCGA TATCAATATT AATACCTAAA	1020
5	TCTTTTTTGA CTTGTGTTAA GATTTGATGA ATAATCGATA AGTTTCTCAA CCCTAGAAAA	1080
	TCTATTTTTTA ATAACCCAAT ACGTTCGGCT TCAGTCATTG TCCATTGCGT TAATAATCCT	1140
	GTATCCCCCTT TCGTTAAAGG GGCATATTCA TATAATGGAT GGTCAATTAAT AATAATTCCT	1200
10	GCCGCATGTG TAGATGTATG TCTTGGTAAA CCTTCTAACT TTTTACAAAT ACTGAACCAG	1260
	CGTTCATGTC GATGGTTTCG ATGTACAAAC TCTTTAAAT CGTCAATTTG ATATGCTTCA	1320
	TCAAGTGTA TTTCTAATTT ATGTGGGATT AACTTGAAA TTTCAATTAA TGTAACCTCA	1380
15	TCAAACCCCA TAATTCTTCC AACATCTCTA GCAACTGCTC TTGCAAGCAG ATGACCGAAA	1440
	GTCACAATTC CAGATACATG TAGCTCGCCA TATTTTCTT GGACGTACTG AATGACCCTT	1500
20	TCTCGGCGTG TATCTTCAA GTCAATATCA ATATCAGGCA TTGTTACACG TTCTGGGTTT	1560
	AAAAACGTT CAAATAATAG ATTGAATTTA ATAGGATCAA TCGTTGTAAT TCCCAATAAA	1620
	TAAGTGACCA GTGAGCCAGC TGAAGAACCA CGACCAGGAC CTACCATCAC ATCATTGCTT	1680
25	TTGCGATAAT GGATTAAATC ACTTACTATT AAGAAATAAT CTTCAAACC CATATTAGTA	1740
	ATAACTTTAT ACTCATATTT CAATCGCTCT AAATAGACGT CATAATTAAG TTCTAATTTT	1800
	TTCAATTGTG TAACTAAGAC ACGCCACAAA TATTTTTTAG CTGATTCATC ATTAGGTGTC	1860
30	TCATATTGAG GAAGTAGAGA TTGATGATAT TTAAATTCTG CATCACACTT TTGAGCTATA	1920
	ACATCAACCT GCGTTAAATA TTCTTGTTA ATATCTAATT GATTAATTTT CTTTTCAGTT	1980
	AAAAATGTG CACCAAAATC TTCTTGATCA TGAATTAAGT CTAATTTTGT ATTGTCTCTA	2040
35	ATAGCTGCTA ATGCAGAAAT CGTATCGGCA TCTTGACGTG TTTGGTAACA AACATTTTGA	2100
	ATCQAAACAT GTTTTCTACC TTGAATCGAA AACTAAGGT GGTCCATATA TGTGTCATTA	2160
	TGGGTTTCAA ACACTTGATC AATATCACGA TGTGATCAC CGACTTTTTT AAAAATGATA	2220
40	ATCATATTGT TAGAAAATCG TTTAATAAT TCAAACGACA CATGTTCTAA TGCATTCAAT	2280
	TTTATTTCCG ATGATAGTTG ATACAAATCT TTAAATCCAT CATTATTTTT AGCTAGAACA	2340
45	ACTGTTTCGA CTGTATTTAA TCCATTTGTC ACATATATTG TCATACCAA AATCGGTTTA	2400
	ATGTTATTTG CTATACATGC ATCATAAAAT TTAGGAAAAC CATACAATAC ATTGGTGTCA	2460
	GTTATGGCAA GTGCATCAAC ATTTTCAGAC ACAGCAAGTC TTACgGCATC TTCTATTTTT	2520
50	AAGCTTGAAT TTAACAAATC ATAAGCCGTA TGAATATTTA AATATGCCAC CATGATTGAA	2580
	TGGCCCCCTT CTATTAGTTA AGTTTTGTGC GTAAAGCTGT AGCAAGTTGC TCAAATTCAT	2640

55

	CAATATCATT AATAATCAAT TGCCCTTTAG AACGTAATCG ACATCTGATT TCATTACCTT	2760
	CATCGACTGC AAATACCCAT ATTTTCAAGC CTTTGTATGTC AGCAATTGTA TTAACAAACT	2820
5	GAGATGCTTC ATTTGGCTGA ATACCGAATT GCTCCAATAC ATCTTCAGTT ATTTTAACTT	2880
	GGCAGAATCC ATCATCCATA AGTTCGAAAT GTTGTAACAA ATAACCTTGA AACGGCAACA	2940
	TTTTTGGGTC CTTCTCCATC ATTTTATTTA AAAGCGCATT ATGATCAATA TCATGCCCAA	3000
10	TTAACTTTCC AGCAATTTCC ATAGTATGTT CTGAGGTATT GTTAAAAAGG AATCGCCCAG	3060
	TATCACCGAC GATACCAAGA TATAAACGC TCGCGATATC TTTATTAACA ATTGCTTCAT	3120
	CATTAAAATG TGAGATTAAA TCGTAAATGA TTTCACTTGT AGATGACGCG TTCGTATTAA	3180
15	CTAAATTAAT ATCACCATAC TGATCAACTG CAGGATGATG ATCTATTTTA ATAAGTTTAC	3240
	GACCTGTACT ATAACGTTCA TCGTCAATTC GTGGAGCATT GGCAGTATCA CATACAATTA	3300
20	CAAGCGCATC TTGATATGTT TTATCATCAA TGTTATCTAA CTCTCCAATA AAACCTTAATG	3360
	ATGATTCCGC TTCACCCACT GCAAATACTT GCTTTTGCGG AAATTTCTGC TGAATATAGT	3420
	ATTTTAAACC AAGTTGTGAA CCATATGCAT CAGGATCTGG TCTAACATGT CTGTGTATAA	3480
25	TAATTGTATC GTTGTCTTCG ATACATTTCA TAATTTCAAT CAAAGTACTA ATCATTTTCA	3540
	TACTCCCTTT TTTAGAAAAG TTGCTTAATT TAAGCATTAG TCTATATCAA AATATCTAAA	3600
	TTATAAAAAT TGTTACTACC ATATTAACT ATTTGCCCGT TTTAATTATT TAGATATATA	3660
30	TATTTTCATA CTATTTAGTT CAGGGGCCCC AACACAGAGA AATTGGACCC CTAATTTCTA	3720
	CAAACAATGC AAGTTGGGGT GGGGCCCAA CGTTTGTGCG AAATCTATCT TATGCCTATT	3780
	TTCTCTGCTA AGTTCCTATA CTTCTGCAAA CATTTGGCAT ATCAGGAGAG CGCTCGCTAC	3840
35	TTTGTGTTT TGACTATGCA TGTTCACTTC TATTTTGGCG AAGTTTCTTC CGACGTCTAG	3900
	TATGCCAAAG CGCACTGTTA TATGTGATTC AATAGGTAAT GTTTTAATAT ACACGATATT	3960
	TAAGTTCTCT ATCATGACAT TACCTTTTTT AAATTTACGC ATTTTCATATT GTATTGTTTC	4020
40	TTCTATAATA CTTACAAATG CCGCTTTACT TACTGTTCCG TAATGATTGA TTAAAAGTGG	4080
	TGAAACTTCT ACTGTAATTC CATCTTGATT CATTTGTATA TATTTGGCGA TTTGATCGTT	4140
45	AATTGTTTCA CCCATCTGAG GCTGTCTTCC TAAAAGTTGC ATAGACTTTA AAACATCTTG	4200
	TCTATTAATC ACACCCACTG TCTTTTATT ACTCGAAACG ACAGGAATCA ATTCAATACC	4260
	TTCCCAAATC ATCATATGCG CACAACTTGC TACTGTACTC ATAGCATTTA CATAAATAGG	4320
50	ATTTTCGCTC ATCACTTTAT CTATTTCTGC GTCGTCCTTT GTATTAATCA TCTCTCGACT	4380
	TGTTACAATA CCTACTAATT TATACGACTC ATTGACTACC GGAAATCTTG TATGGCCAGT	4440

55

ATCTAATGGC GTCATTATAT CTTGAACTAT TAAGATATCT TTTCGTATTT TCTGATTAAA 4560
 AAGTGCTTTG TTGATAATAT TTGCAACTAG GAATGTATCA TAACCTGATG ATAGAACAGG 4620
 5 TAAATCATGT TCATTCGCAA AATTAATAAC TTTATTAGAT GGCTTAAATC CACCAGTAAT 4680
 TAATATAGCC GTACCTCTTT TTAAGCTTC AATCTGCACA TCTTCACGAT TTCCGACAAT 4740
 CAATAATGTC TTGGACCAA TATACTTTAA AATATCTTTG AGTTCCATTG CTCCAATTGC 4800
 10 AAATTTAGAT ACCATCTTAG TGATACCTTT GTTGCCACCT AACACTTGGC CATCAATAAT 4860
 ATTGACAATT TCATTAAAAG TTAAATGTTC AATTTCAATTA CGATTACGTT TTTCGATTTCG 4920
 AACCGTACCA ACACGATCTA TCGTTGCGAC CATGCCCATTT TTATCAGCAT CTTTmATTGc 4980
 15 ACGATATGCT GTCCCYtCaG ATACGTTTAA AAATTTAGCG ATTTTACGCA CCGAAATTTT 5040
 AGAGCCTATA GATAACGATT CAATATAATC TAAAATTTGT TCATGTTTTG TCATTCTTTA 5100
 CCTCTTCTTT TCGAACAGTA TTAACCTACAT TATAACTTTA TTTTGGATAA AAAGCATTGA 5160
 20 AGTGAAATGA AATAATGATC GTTtCACCTA TTTTATTTTT TGAAAATATA CAACAAACAC 5220
 AAAGATCACA AAATCTTTAA TTTTAAATGG AAAATCCAT TATTATTTAT TAGAATGTAA 5280
 GTGAGGAGGG ATGTACTAAT GTATAAAAT ATATTACTTG GTGTAGACAC TCAGTTAAAA 5340
 AATGAAAAAG CACTAAAAGA AGTGTCTAAA TTAGCTGGCG AAGGTACAGT CGTAACAGTT 5400
 TTAAACGCAA TCAGCGAACA AGaTGCTCAA GCATCAATTA AAGCAGGTGT TCATTTAAAC 5460
 30 AAACCTTACTG AAGAACGAAG CAAGCGATTG GAAAAACAC GCAAAGCTTT AGAAGATTAT 5520
 GGTATTGATT ATGACCAAAT AATTGTTTCGT GGTAAATGCAA AAGAAGAACT ATTAAAAAT 5580
 GCTAATAGCG GTAAATATGA AATTGTTGTT TTAAGTAACC GTAAAGCAGA AGACAAAAAG 5640
 35 AAATTTGTAC TTGGAAGTGT CAGCCACAAA GTAGCAAAAC GTGCGACTAT CCCTGTATTA 5700
 ATCGTTAAAT AAAATTTTTA TCCAGAATCA CAAATAATCT TTCAATCATG ATGCAGTCTC 5760
 AAACGACTGA GTAAATACAA GAAACGATTA TGACTGTGGT TCTGGATTTT TTATATCGTA 5820
 40 GTAAATTTAT AATCAATGTC TAATTGTATA AACTAAAAT TACGAGAGTA GGTCAGAAAT 5880
 GATAAAGAAC CACTGATGTC CCCCCTCCAC GTCGTAACCTG AATCAGTAGA ATATAAAAAC 5940
 ACCCACTAAA AATATGCAGA CGATAACTTC CACATAGATT AGCGAGGTGT TTTTtagTGT 6000
 45 AAAATCTATA TTCTATTTAA AACTGAACAG ATTCACCTGG TTTTAAATTT TGCACGTCCC 6060
 CTACATTAAAC AGCATCTTTA AATTGTTGTG GATCTGTGTC GATTAAATGGG AATGTATCAT 6120
 50 AATGAATCGG TACAGAAATT TTTGGTTTAA TAAATTCATT AATAGCATAA CTTGCATCAT 6180
 CAATACCCAT CGTAAATTA TCTCCAATTG GTACAAAACA TACATCAACT GGATGACGTT 6240

55

	TTCAACTTCA AACACGATAC CCATTGGCAT ACCTAAATAA ACTGGgAATA CCATTTTCAT	6360
	GTGTAAAACT TGAACATGA AATGCTTGAA CAAATTTAAC GCTTCCGAAA TCAAAgTTTG	6420
5	CTTTACCACC AaTATTCATA CCATGAACAT TTTCAACACC GTGATATGAA GAAAGATAGT	6480
	CAGCCATTTT TGCACCTCCA ATTACTGTTG CTCCTGTTTT CTTTGCTAGT TCCACAACAT	6540
	CACCAAAATG ATCAAAATGA CCGTGCGTTA AAACGATATA GTCTACCTGC ACTGTTTCAA	6600
10	TATTCAAATC AACTTAGGG TTATTTGAAA TAAACGGATC TACGATAACC TTTTGTGTGT	6660
	TCCCTTCTAA ATAAATCGTT GATTGACCAT GAAATGATAA CTTCAATTGA GCATCCTCCT	6720
	ATCAATTACT ATATAAATTT AGTACCCTTT TGCCACTTAA TTATAACAAA TTCTCAAATT	6780
15	TTAAAAATTG AAAATCTAGT TAATGTATTA GCTCGATTTT GAAATCTAAT AATAATTGGC	6840
	ATAAAATGGA AGTAATATTA TGTGAGGAG TGTTTATAAA ATGACAAAAA TATCAAAAAT	6900
20	AATAGACGAA TTGAACAATC AACAAAGCTGA TGCAGCATGG ATTACAACAC CGTTGAATGT	6960
	ATATTATTTT ACTGGATACC GTAGCGAACC CCATGAAAGA TTATTTGCAT TATTGATTAA	7020
	GAAAGATGGT AAACAAGTAC TATTTTGTC CAAAATGGAA GTCGAAGAAG TCAAAGCATC	7080
25	ACCTTTCACA GGTGAAATCG TTGGATATTT AGACACTGAA AACCCTTTTT CACTTTATCC	7140
	TCAAACAATC AATAAATTAC TAATTGAAAG CGAGCACTTA ACAGTAGCAC GCCAAAAACA	7200
	ATTAATCTCT GGTTCATG TCAATTCATT CGGAGATGTT GATTTAACAA TCAAACAATT	7260
30	GAGAAATATT AAATCCGAAG ATGAAATTAG CAAAATACGT AAAGCTGCTG AGTTAGCAGA	7320
	TAAGTGATC GAAATAGGTG TTTCTTATTT AAAAGAAGGT GTGACTGAAT GTGAAGTAGT	7380
	CAACCATATT GAGCAAATA TCAAACAATA TGGCGTCAAT GAAATGAGTT TTGATACGAT	7440
35	GGTTTTATTT GGAGATCATG CCGCATCACC TCATGGCACA CCAGGAGATC GCAGATTAAA	7500
	AAGCAATGAA TATGTACTAT TTGATTTAGG TGTAATTTAT GAGCATTATT GTAGCGATAT	7560
	GACACGTACT ATTAAATTTG GTGAACCTAG CAAAGAAGCA CAAGAAATTT ATAATATTGT	7620
40	ATTAGAAGCA GAAACATCTG CAATCCAAGC AATTAAACCT GGAATACCAT TAAAAGATAT	7680
	CGATCATATC GCTAGAAATA TTATTTTCTG AAAAGGTTAT GGTGAATATT TCCCTCATCG	7740
45	CTTAGGTCAT GGCCTAGGAT TACAAGAACA TGAATATCAA GATGTTTCAA GTACTAATTC	7800
	TAATTTGTTA GAAGCTGGCA TGGTTATTAC AATCGAACCA GGTATTTATG TACCTGGTGT	7860
	TGCAGGTGTA AGAATTGAAG ATGACATACT TGTCATAAT GAAGGATATG AAGTATTAAC	7920
50	ACATTACGAA AAATAAGGAG TGGGATAAAA ATGAAAAGCT TGTTACAAGC GCATTCTCAT	7980
	TCAGTCAAAC ACTGCCAATA TAACATTGTA GCGCCTAAGA CATAAATTTT TATCCAAGTC	8040

	TGTAATGAAT CAAATCAATA TCATTCATGT TCGATGATT CTTCGCATTG TTTCTAGCTT	8160
	TAATTTATCA TTATTTAATT TTAATAACCA AGGAGATGAT AACGTCATT CTTAGTACGC	8220
5	TGTAATCCAT TCCCTTTTCA TCAAATTCAA ATTATAATTG TAATGCTTCT TCTACAGATT	8280
	TATATTCCAT TTCAAATGCC TCTGCAACGC CTTTATTGGT TACGTGACCT TTGTAAGTAT	8340
	TTAAACCTAA TGATAATGGT TGATTTGATT TAAATGCTTC TCTATACCCT TTATTAGCTA	8400
10	GCATGAGCGC ATAAGGTAGC GTAGCATTAT TTAAAGCTAA CGTCGAAGTA CGCGGTACTG	8460
	CACCTGGCAT ATTTGCAACT GCATAATGAA CCACACCATG CTTAATATAT GTAGGATCAT	8520
	CATGTGTCGT AATTTTATCA GTTGtTTCAA AAATACCGCC TTGATCAATA GCAATGTCAA	8580
15	TAATAACTGA CCCATTTTTT ATTTGTTTAA TCATGTCTTC TGTTACAAGT CTTGGCGCTT	8640
	TAGCACCTGG AATTAAAACT GCACCTATTA CTAAATCACT TTGTTTAACTA TACAACTCAA	8700
	TATTCAACGG ATTTGACATA ATTGTATGTA CACGTCCACC GAATAAATCA TCTAATTGTT	8760
20	GTAACGCTT TGGATTAACA TCTAAAATCG TAACATCTGC ACCTAGTCCT AGTGCAATTT	8820
	TAGCTGCATT TGTTCTGCT TGACCACCAC CGATAATAGT TACTTTACCC TTAGGTACTC	8880
25	CTGGGACACC ACCTAGTAGA ATTCCCATAC CACCATTAAAG TTTTGTAGG AACTCTGCGC	8940
	CAACTTGAGC TGACATTCTT CCTGCTACCT CACTCATTGG TGATAACAAT GGTAAAGATC	9000
	GGTCTGGTAA CTGCACAGTC TCATATGCAA TACTAATTAC TTTTCTATCT ATCAAAGCTT	9060
30	GTGTTAATTT TTCTTCATTT GCTAAATGaa gatAaGTGAA TAATACAAGC CCTTCTTTAA	9120
	AATATGGATA TTCAGATTCA AGTGGTTCTT TAACTTTAAT AACCATATCC ACATCCCAAA	9180
	CTTTTGCTTG TTCAGCAACA ATCTCAGCAC CTGCTTCTTT GTAATCTACA TCTTCAAAGA	9240
35	ATGATCCTGA ACCCGcATTT GTTCCACTA AAACAGTATG	9280

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:

- | | |
|----|-----------------------------|
| 40 | (A) LENGTH: 4669 base pairs |
| | (B) TYPE: nucleic acid |
| | (C) STRANDEDNESS: double |
| | (D) TOPOLOGY: linear |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

	CTGATTAATC TCTTGTGTGTC GTGTATTTAC TAATTGAATC GTTGGTGTCT GAACACGTCC	60
50	CAGGGATAGC TGTGCATCAT ACTTGTGTGT TAGTGCACGC GTTGCAATTAA TCCCAACAAT	120
	CCAATCTGCC TCACTTCTCG CTAACGCTGC ATAATACAAA TCGTTATATT GACGACCGTC	180

55

	ACGGATTGGC TTTTGTAC CAACTTTATC CAAAATCAAT CTTGCAACTA GTTCACCTTC	300
	TCGTCCaGCA TCTGTTGCAA TAATAATATC TTTCACTTTA TTATCTAAAA TTAACGCTTT	360
5	TACTGTTTTA AATTGTTTGC TTGTTTACC AATAACAACA GTTTTCATAT ATTTAGGTAT	420
	AATTGGAAGG TCTTCTAATC GCCATTCCCTT TAAATTTTTA TCGTATTGTT CAGGTGTCGC	480
	ATTTGTCACT AGATGACCTA ACGCCACGT GACAATATAT TGGTTATTTT CAAAGTAACC	540
10	ATTACGCTTC TGATTTATTT GTAAAGCATC AGCAATATCT CTTGCGACTG ATGGTTTTTC	600
	AGCTAATATT AAAGATTCA TAAATTATCC TTTCTCATAC GTTCTTTTAT TTCGAACGTG	660
	CTTCATCTAT TCCACTAATC TTTGATTTAA ATTCAATGAT TGCAAATGAT GTGTTAAATG	720
15	TATTGTAACA TGTTAATATC ACTATTAAC TTCATTTTCA GTGAAATACT ATATAATAAA	780
	AGTAACAAAA AGTACGGAGG TAATGACATG AGCATAGTTC AGTTATATGA TATTACACAA	840
20	ATAAAATCGT TCATTGAACA TTCGAATTAT GAATCAGCAT CATACTTATA TAAACTTCCT	900
	CAACAGTACA ATGAAATAGA TGTATTAATA ACCGATGCGA TTGAATCACC TGGTGTATTT	960
	TCGATTAAAG AAAACGATTC AATCAAAGCA ATCATATTGT CTTTTGCATA CGATAAAAAT	1020
25	AAATTCAAAG TCATAGGCCC TTTCGTGGCT GACAATTATG TATTATCTGT CGATACGTTT	1080
	GAAACGCTAT TTAAAGCAAT GACTTCGAAC CAACCTGACG ATGCCGTCTT TAACTTTCT	1140
	TTTGAAGAAG GCATTCAACA ATACAAACCA TTAATGAAAG TTATTCAAGC AAGTTATAAC	1200
30	TTCACTGACT ATTACATAGA AGCCCGTACA AGATTAGAAG AAGATATGCA CCAACCAAAT	1260
	ATCATTCCCTT ATCACAAAGG GTTTTATCGT GCTTTCAGCA AATTACACAC AACTACATTT	1320
	AAATATCAGG CACAGTCACC ACAAGATATC ATTGATAGTT TAGACGACCA TCATCATTTG	1380
35	TTTTTATTTG TTAGCGAAGG TTTACTTAAA GGTTATTTAT ACCTTGAAAT TGATTCACAA	1440
	CAGTCAATCG CCGAGATTAA ATACTTCAGT TCTCATGTAG ATTACCGTTT GAAAGGTATC	1500
	GCTTTCGAGT TGCTTGCCTA TGCATTGCAA TATGCTTTTG ATAATTTTGA TATTAGAAAA	1560
40	GTTTATTTTA AAATTCGTAA TAAAAATAAT AAATCATCG AACGATTTAA TGGTCTAGGT	1620
	TTCCATATCA ACTATGAGTA CATTAAATTC AAATTCGAAT CACGTAACGT AAAAGATCAA	1680
45	ACAATCCCTG AATAAACAC CAAGCAAATA CCCTACAGTA CATCATTAGC ATGTATTGTG	1740
	GGTTTTTCTA CTTTTGTAA ATATTGAAAA TTATAAGTAG TTGTTTTTTA CTATTAGGGC	1800
	AGAATGCTTT ACAATAACAT GCAAGTGTC AATTAAGGGA GCACTTGCAT AAATAGTATA	1860
50	GGAGAGTGAG TAGTCTTGCA ATTTCTTGAT TTCTTAATCG CACTTTTACC TGCTTTATTC	1920
	TGGGGAAGTG TCGTTCCTAT TAATGTGTTT GTCGGCGGTG GACCTTACAA CCAAATTCGT	1980

55

	TTCAATAATC	CTACTGTAAT	TATTGTCGGT	CTTATTTCTG	GTGCATTATG	GGCGTTTGGG	2100
	CAAGCGAATC	AGCTTAAATC	TATTAGTTTA	ATCGGTGTAT	CAAATACTAT	GCCAGTTTCT	2160
5	ACAGGTATGC	AATTAGTTGG	TACAACATTA	TTCAGCGTTA	TCTTTTTAGG	TGAATGGTCT	2220
	TCAATGACTC	AAATTATCTT	TGGTTTAATC	GCCATGATAT	TATTAGTTAC	TGGTGTAGCA	2280
	CTTACTTCAC	TTAAAGCTAA	AAATGAACGT	CAATCAGATA	ATCCTGAATT	TAAAAAAGCA	2340
10	ATGGGTATTT	TAATTGTATC	TACAGTTGGA	TATGTAGGTT	TCGTTGTACT	TGGTGACATC	2400
	TTTGGTGTG	GTGGAACGA	TGCATTGTTT	TTCCAATCTG	TCGGTATGGC	AATTGGTGGC	2460
	TTTATCCTAT	CCATGAATCA	TAAAAACATCA	CTTAAATCAA	CAGCACTTAA	TCTATTGcCA	2520
15	GGTGTGATTT	GGGGAATTGG	TAACTTGTTT	ATGTTCTATT	CTCAACCAAA	AGTTGGTGTG	2580
	GCTACAAGTT	TCTCATATATC	ACAGTTACTT	GTTATCGTTT	CAACCTTAGG	CGGTATTTTC	2640
20	ATTTTAGGAG	AAAGAAAAGA	TCGTCGTCAG	ATGACGGGTA	TTGGGCAGG	TATTATTATT	2700
	ATCGTGATAG	CTGCTATAAT	TCTAGGTAAT	TTGAAATAGA	AAGTTAAATA	CTCATGTAAC	2760
	GTAAAAATGT	AATCACTTCT	GAAAATAACC	ATCACTTAT	AGAATGATTA	AAATTAATTT	2820
25	TCGGGAATTT	TACGTTGAAT	GTTCTCTAT	ATGTCCTAGG	AAATACGTGG	CTCTAAAAAC	2880
	AAAACGCAAT	AACACATCAT	GACATTAATC	ATGCGTTTTA	AGACTTTAAA	ATTAGCGATA	2940
	CTTTTAAAAT	CTTGATGATA	TTCATATATC	AAGTATGCGC	CATACATATG	AAGTGGATAG	3000
30	CTGCATAACG	CACTGCATTA	TCAACTTGAA	TGTATGAGTT	GAACAACTAT	GTCATAAATA	3060
	AAAGCCCCCT	TTTACAATA	TACATTTACA	TATTGTGGTA	AAGGGGGCTC	TCATTTTCTA	3120
	CGAATACTAA	AATGGATTTT	ATTTTCAAAT	GTGTAAACTA	GACAAACACT	GCCTGATACA	3180
35	CGTACAAAAT	AATGATACTA	ATAATGATTG	TCAAATTGGT	CGTCATACCT	ATAAATGGCA	3240
	GTGTTTCGATA	TTTAAACTGA	ATACCATAAG	AAATAATTGC	AACACcTACC	GGGAACATCC	3300
40	AAGTGACCAA	CAATGTCGTC	TTAATCATAT	CATCTGATAC	TGGTAACAAG	ACATATACTA	3360
	ACAATCCCGC	AACATATGCT	AATCCATAAT	GCAAACATAA	ATATTTAATA	GTAGCAGGTA	3420
	TATACTTTCT	TTCCAGAGTA	AAATTCAACA	TGACACCTAG	CAAAATCATT	GATAACGGCA	3480
45	TATTTGCATG	GGAAAGTATG	CTAAAGAAAT	CGATTGCCAC	ATGTGGTAAA	TGGATGTGAC	3540
	TTATATTCAA	TATAAACATT	ACAATGTATG	TAACGAGTGG	CACTGATTGT	AATAATTTCT	3600
	TACCTAAATA	TTTAAAATCG	AATTGATCAC	TACCTTCACT	AAAGTAGCTA	CCTACAAAGT	3660
50	AAGTAATTCC	AAACATCACA	AAGGCACCAC	CTATATCAGC	CATAACAAAA	TAAATAAGTC	3720
	CCGTTTTAGG	CCATATCACT	TCAATTAGTG	GATATGCAAA	CAATCCAATA	TTCATAGCAC	3780

55

CAATCATTTT CGCCACAATA CCATATATAA TCATTAAAAAT TGGTAAAAATG GAGAATGACA 3900
 ATTTTAATTC TGCAGTGTTC AAATTCACAA TAACTAAAGA TGGGAGTGTG ACATTAAGAA 3960
 5 CTAATGTAGC AATGACTTGA CTATCTGTTG CTTTTATAAA ATTAATGCGC TTCAAAAAGT 4020
 AACCAAGCGC AATTAATAAA ATAATCATAG TAAATTGTTC TGTCAGTGTG ATCCCTTCTT 4080
 TCAATAATCT TCATAATTTA TAACTTTAAC ATACTCCACA GATATTTTAG AAGTCTACTG 4140
 10 TTTCATGCTA TAATCTACAT TAAATGCACT TAATTATATT TCAAAGGAGT GTTATAGTAT 4200
 GTCTTTAGAA AACCAACTAG CCGAAGCTAA ATATGATTAT GTTCGTCTTC AAGGTGACAT 4260
 AGAAAAACGG GAATCTTTGA ATTTAGATAC TTCCGCACTT GTTCGTCAAC TTAAAGATAT 4320
 15 TGAATAATGAA ATTAGAAACG TTCGTGCTCA AATGCAAGAT TAATAATCTA TCATTCAAGC 4380
 AATAAATGCT TTTTGTTACA TAAATTTGAC TAGCATTGCT CTGAATACGT TATATTGATG 4440
 AATTGCTTCA TTTTCGCTC AATTACATCT AGAATCACAA GATGTTGTGCG TGTATGATT 4500
 20 TAGTGTTC AATAACAACAT ACACGCATAT CTATCCCAAC ACTGCTATTT ATGTTTTCTA 4560
 CGCTGnTGTA CTACATGAAC CCTTTGAAAC GGAGAGGAAG TTATCATATG CAATTTTAnC 4620
 25 TGATTTTACT AGCAATACTT TAACnAATTG nTAGTTTAAT AGAATTTTA 4669

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 2785 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

TTTGcACCCA TCTGaTACAA TGCACCATGC GGTTTAACAT GATTAATTTT AACTTGATGA 60
 ATGCGACAAA ACCCTTGTA TGCACCTAAT TGATAAATCA TCAAATTATA AATCTCGTCG 120
 40 TTAGAGATAT CTATATTTG TCTGCCAAAG CCTTTCAAAT CAGGTAAACC AGGATGTGCA 180
 CCTACTGCAA CATTATGTGC TTTGGCAAGT TTTACGTTT CATTCATTAC ATTTTCATCA 240
 CCAGCGTGAA AACCAAGC AACATTGCA CTGTAAATTA ACGGAATAAT TTGATGATCA 300
 45 CCACCAAAGG AATAATTTCC AAATGCTTCG CCTAAATCAC AATTCAAATC AACTCGCATT 360
 ATAATTCCAC CCCTTTAACA ATTTGATGTT TTTCTAAAAA TTTAATATCA ACATCTTTTG 420
 50 CATCTCCATC ACGATATAGT GGATAATTTA AACTGCATA TAAAAAATCG GCAGTTGTAG 480
 AAAATCCATC TATCACCATT TCATCTAAGG TGACTTTCAA CTTATCAATT GCTGAAGCTC 540

55

	AACCGTGATA TAGTAAAGAA TCGACTCGCA CATTAAAGCC TTGAGGTAAA TGTAACGCTG	660
	TCACTTTACC TGGTGTGGT TGAAATTCT TTTCaGGATT TTCGGCATT ATTCTCGCTT	720
5	CTATCACATG ACCATTAAAT TGAATATCGC TTTGTGAAAA AGGTAAATGA TTATGTTCCA	780
	ATAAATACAG TTGTGCTGCA ACCAAATCAC GTTCTGCTCG CATCTCTGTA ACAGTATGTT	840
	CAACTTGTAT TCGAGCATT CATTCAATAA AGTAATGTGC GGTATCAGTT ACTAAAAATT	900
10	CAATCGTACC TGCACCTCTA TAATTTGCTG CACGTGCAAC TTTAACAGCA TCGTTACATA	960
	TTTGTGTGCG TCTTTCTTCA GTTAATGCTG CACAAGGAGA TTCTTCGATT AATTTTGTGAT	1020
	TTTTACGTTG TACAGAACAA TCACGTTCCC CTAAATGTAC ATAATTATCC TGCCCATCTC	1080
15	CCaTAACITG AACTTCAACA TGTTTTGcAA CAGGTATAAA AGCCTCAACA TAAACACGAT	1140
	CATCATCAA GTATTTTTT CCTTCACTTT TAGCTTCTTT AAATGCCTTT TCTAAATCTT	1200
20	CAGCTTTCTT TACAATACGT ATACCTTTAC CACCACCGCC ACTGGCAGCT TTGATAACAA	1260
	CTGGATAACC GATGTCTTTG GCAAGATTCT CAATTTGAGA CACATGATTC ACAGCACCAT	1320
	TTGATCCTGG AATCACAGGA ACACCTGCAT GATGAAGTGT TTGTCTTGCT GTTATTTTAT	1380
25	CCCCATCAT TTCCATCGTT TTTTATAGTAG GCCCTATAAA CGCTATGCCT TGTTCCTCAA	1440
	CGGTTTGAGC AAATTTTGTT GATTCTGATA AAAAGCCATA TCCTGGGTGA ATTGCATTAG	1500
	CACCAGTGAT TTGTGCAGCA GATATGATGC GGTCAATATT TAAATAACTA TCTAAAgCAT	1560
30	TArCwTCCCC AATACATATA GCTTGATCTG CTAAATGTAC ATGCAAGCTT TGCTCGTCCC	1620
	CTTTTGATA AACTGCTACA GTTTCAATCC CATATCTCT GCAAGCTCTT ATAATCCTTA	1680
	CAGCAATTC ACCTCTGTT GCAATTAAAC AACGAAGCAT TTAATTACCC CCTTTACTTA	1740
35	ATACGTACCA AAAGTTGGTC GTATTCAACA TTTGTGCCAT GATCAGCTAC TATTTAGTA	1800
	ATTCTCCAG CAACATCTGT TGTTACCTCG TTTAATACTT TCATCGCTTC AACATATCCT	1860
	ATAATATCTC CCTTGTTAAC TTTGTCACCG ACATTCACAA TTGGTTCAGT TAATTCTTTA	1920
40	CTATCTTGTA AAAAGAATGT ACCTATCATT GGTGATTTAA TGTCATGATA ATCATTGTGTC	1980
	GAAACATCGG AGTTATCATT CGCTTTTGAA GCTGTCAAAT CATTATTGTT CATACTTTGA	2040
45	TTTGATTGAT TACTGTGTGC AGCCAAATGA TTCGAGTCAG TGAAGTCAAT TTCTATTTCA	2100
	TCTTCAAAAT TTTTATATTT AAATTTCTTA ACATCATTTT CCTTCACTAA TTTGATTATT	2160
	TGTTTCGATTT nTTCAATATT CATTTTACAA ATCCCCTTTT AAAATTGTTG CTAATTTTTT	2220
50	CGAAGTATGT CGCAAGCTAG ATGTATCAAA AATTGGAGTC TTTTGATGAC TCTTAAGAAT	2280
	TTCATTAAAC AGAGACATTT GTTCCCGATT CTTATCTACA GCTTCTTGGA ATGATATCCA	2340

55

TACAGTTGCA ATTTTGGTAT AACCACTAT CGTTTGTTTA TCATTAAGCA GAATAATAGG 2460
 TTGACCATCA TTTGGTACCT GAACACTACC AAGAGCAACC GGTTTCAGAAA TGATATCTGC 2520
 5 TTGATTAACT GGTGCAACGC TGTCACCTTC CAAACGATAG CCCATACGGT CTGATTGTTC 2580
 AGTAATTAAA TATGGATGAT TTACAATTTT CGCTCTAGCC TCTTCAGAAA ATGCCTCGAA 2640
 TTGAGGTCCT TGAAGAATGT GTATAATATT ATTTTCTGGC AATAAATCGT CCTGTAAATG 2700
 10 AATCGTCTTT CCAATGTTTT CTTTAAAGTC ATTATTTATT TTCACTGTTA TTACATCATC 2760
 AGCTAATAAC TTTCTACCTT TGAAT 2785

(2) INFORMATION FOR SEQ ID NO: 134:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

25 AATGGAAACG GTTGAACAG CAATTATTAC TATTTCTATG GGTGAAGGTA TTTCAGAGAT 60
 ATTTAAATCA ATGGGTGCCA CACATATCAT TAGTGGTGGA CAAACGATGA ATCCTTCTAC 120
 AGAAGATATC GTTAAAGTCA TTGAACAATC AAAATGTAAA CGTGCAATTA TTTTACCGAA 180
 30 TAATAAAAT ATCTTAATGG CAAGTGAACA AGCAGCGAGT ATTGTTGATG CAGAAGCTGT 240
 TGTATTCCCA ACGAAATCTA TTCCTCAAGG TATAAGCGCA CTATTCCAAT ATGATGTGGA 300
 CGCAACACTT GAAGAAAATA AAGCGCAAAT GGCTGATTCA GTAAATAACG TTAAATCTGG 360
 35 TTCATTAACG TACGCTGTTC GTGATACGAA AATTGATGGC GTTGAGATTA AAAAAGACGC 420
 GTTTATGGGC TTGATTGAAG ATAAGATTGT AAGCAGCCAA AGTGATCAAT TAACAACGGT 480
 TACTGAGTTG TTAAATGAGA TGTTAGCAGA AGATAGTGAA ATATTGACTG TGATTATTGG 540
 40 TCAAGATGCA GAGCAAGCAG TTACAGATAA CATGATAAAC TGGATCGAAG AGCAATATCC 600
 AGATGTAGAA GTGGAAGTTC ATGAAGGTGG ACAACCAATT TATCAATATT TCTTTTCAGT 660
 AGAATAAAAA TTAAATAA AAAACTACCA ATGATAAATC ATCAGTTGGT AGTTTTTTAT 720
 TTTGCTATTT TAGTGATATT GCGGGTTAAA AGTATCGTTC TCGAGTTGCT AACAATGTCA 780
 TGTTCAACTT AGTCATGATA AAATAAATAA CATACTAAAT GATACGTAAA ATCAAATAAA 840
 50 ACATAGGTGA TTTATTTTGG CTAAAGTAAA CTTAATAGAA AGTCCATATT CTCTTTTACA 900
 ATTAAAAGGT ATAGGTCCTA AGAAAATAGA AGTATTGCAA CAACTAAATA TTCATACAGT 960

55

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1540 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

TGTAGTTGAA	CATGAACAAC	AAAAGAAAGA	AAAGACAAAA	AAGCAATACA	AGCCATTTTG	60
GATTGTCATG	AGTTTATATA	TACTTATAGT	TGTACTATTA	CTCCCGGCAC	CTTCAAGTCT	120
GCCGATAATG	GCTAAGGCAG	TACTAGCTAT	TTWAGCTTTT	GCAGTTATTA	TGTGGGTAAC	180
GGAAGCTGTA	TCATATCCGG	TGTCAGCAAC	TTTAATTATT	GGCTTAATGA	TATTACTTTT	240
AGGATTTAGC	CCTGTTCAAA	ATTTAGGGGA	GAAGCTAGGT	AATCCGAAAA	GTGGCAGTGC	300
TATTTTAGCT	GGAAGTGACC	TTCTAGGAAC	TAATCATGCA	TTATCATTAG	CGTTTAGTGG	360
ATTTGCAACT	TCAGCTGTAG	CTCTCGTTGC	AGCTGCATTA	TTTTTGGCTG	CTGCTATGCA	420
AGAAACGAAT	TTGCATAAAA	GACTAGCTCT	TTTAGTGTTA	TCAATTGTTG	GTAATAAAAC	480
TAGAAATATA	GTTATTGGAG	CAATTATCGT	TTCAATTGTA	CTTGCATTTT	TCGTTCCCTC	540
TGCAACAGCT	AGAGCAGGGG	CAGTTGTACC	AATCTTGCTG	GGTATGATTG	CGGCATTTAA	600
AGTTTCCAAA	GATAGCAAGT	TAGCGTCTTT	ATTAATAATT	ACTTCAGTAC	AAGCTGTGTC	660
AATTTGGAAT	ATTGGTATCA	AAACGGCGGC	AGCACAAAAT	ATCGTAGCGA	TTAATTTTAT	720
AAACCATCAA	TTAGGATTTG	ATGTTTCATG	GGGCGAGTGG	TTCTTATATG	CAGCGCCTTG	780
GTCCATAGTT	ATGTCCGTAG	CTTTATATTT	CATCATGATT	AAAGTGATGC	CTCCAGAAAT	840
TAATACAATA	GAAGGTGGTA	AAGATTTAAT	AAAAGAAGAA	TTGCATAAAC	TTGGCCCCGT	900
TAGCCACAGT	GAATGGCGTT	TAATTGTTAT	ATCGATGTTA	TTATTACTGT	TTTGGTCAAC	960
TGAAAAAGTA	TTACATCCGA	TTGACTCTGC	ATCCATTACT	ATTATTGCTT	TAGGTGTTAT	1020
GTTAATGCCG	AAAATTGGTG	TCATGACATG	GAAACATGTT	GAAAATAAAA	TACCATGGGG	1080
AACAATTATC	TGTTTTGGTG	TAGGTATTTT	ACTAGGTAAC	GTTCTTTTGA	AAACAGGTGC	1140
AGCTCAATGG	TTAAGTGATC	AAACTTTTGG	TGTTTTAGGT	TTAAACATT	TACCTATTAT	1200
CGCGACAATT	GCACTTATCA	CGCTTTTTTA	TATATTGATT	CATTGGGGCT	TTGCGAGTGC	1260
AACAAGTTTA	TCATCAGCGT	TAATACCTGT	TTTTATTTCG	CTAACCTCTA	CGTTACACTT	1320
AGGAGACCAG	TCTATAGGAT	TTGTTTTAAT	TCAACAATTT	GTTATTAGTT	TTGGTTTCTT	1380

AGATTTCTTG AAGGCAGGTA TACCATTGAC AATTGTAGGG aATAtctAgT GaTAGTTTTT 1500
 AGCATGACTT ATTGGAAATG GGTAAGGTTG CnTTAATTAA 1540

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11823 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

ACTTCTCACA ATAAGAAATA TGAAATTGTT ATGTGTTAGT TGAGATTCAG TGATGAATTA 60
 CTTTTATCAT TTAAATGTT GTTATCATTG TCATGCGTTA CCAAATCGCT TACGTATACA 120
 CGATTCCCAA TCTTAACATA GACGATTTGT ATATCAGAAT TTTCTGATTA CTAACAGTTT 180
 ACCTAAGTTT AAATATCTGT TCAATGATTT TCAGTTATTT TTAAAAGAAA AATCGTAATG 240
 CTGCCATGAT AACAATCCCA CTAATAATTG TAATAGTTAA AtACGCGTGA TTATAGATAA 300
 AATAACCGTC GGAATGAGCG CGATAATGTA AGGGATGTTT AATGTATACC CCTCACCATG 360
 AGGCGTCTGT TGAATAATGC TGTCAATGAC AAGTGCCGTA AATAGTGTGA TTGGGATAAA 420
 TGATAGCCAT CGAACCACGA CATCAGGCAA TTGCACTTTT GAAATCATGA TAAAAGGTAT 480
 AATTCGAATT AATAGCGTTA CGATACCACA CAATAAAATA AGTATTAACA TGTTCATATG 540
 AGTTATCATT GTTCCATCAT CACTCCTAAC GCTGCTGAAA TTGTGGCTGC AATTAATATT 600
 GCTAGATATG AAGGCATAAA CATACTTAGC GATAACATCA TTACTATGAC GGCAATAATG 660
 AGTACTATGT AAATTCTTAA TCGCGATTTA GTAATTGATT CAAATTGCGC AATGGCCAAA 720
 AAGATAAACA TAGCCGTGAT AGCAAAATCT AACCCTAGCG TTTGCGGATT TGAGATATAT 780
 TCGCCAAATA AAGCCCCAGC TACACATGAA ATTGCCCAAA ATAAATATGC TGTGATGTTA 840
 AGACCATGCA TCCAACGATC ATTGATAGCT TCTCCTTTTA AATAAGGTGT AATGGCGACG 900
 CCAAACGTTT CGTCAGTTAC TAATGAACCT AATCCAACAC GGTTCCAAAA CCCATATGTC 960
 TTGAAGTTTG GTGCAAGCGA CATACTTAAA AGGAACATTC TTGAATTTAC GATAAATACA 1020
 GTTAGTACAA TCGCTGATAT AGGTGTACCT GCTATAAACA ACGCGCACAT AATAAATTGC 1080
 GCAGcaCCGG CATATATAAC AAGACATAAC AAGACAATTT CTAAAATACT AAAGTTTTGA 1140
 GACGAAGCCA CAATACCAAA TGAAATACCA ACACCGGCAT AACCCAATAA TGTGGGATA 1200
 CACTCTTGCA CGCCTTGCT AAAACTTAAA TGTGTTGTCA TCTCAATTAC CTCCTTTGCC 1260

	TAAGCAATAA CATTAGACAT CAGTTTGTCT GAGGTTAGAC ATTCCGGAGT CTTTAGTCAG	1380
	CTTCATATTA ACTTTTTATT TTTGAGAATT TTCAATTTTT TATTTAAGAC TACCTCCATA	1440
5	TTTTCTATGG aTTTGTAGTT GTTTTAAAGT ATCAATTTTA TAAATTTTAA TATCTGATGA	1500
	TGAGTCTGGG aTATTGaTTC ATGTACCACT CCCTTaTaAT CATCCCCTCC CCCTaCCCTA	1560
	CTCCATCGAT ATAACTCATA CTACATATCA ACGAAATCAG TATTTTATCG CTTCCTTTCC	1620
10	TATATTAGTG ATGCTCAAAC TTGTTACGTT TTAGATTGTT TTAGTTCATC ATAATTATCC	1680
	CGTATTGTTG CTATAATGAA ATGCGTTCAC CCCATTAAAC CACAACTTA ATTTATTGTT	1740
15	GTTATGTGCA TTGGCTCACT ATTATATTTT TACAGCACAA AAAAAGTGGC GACAGTTCGT	1800
	CACCACTTTT TAAAATATTA TTTAAAGTAT CTTGCCCTTG CTTTAAGTAT ACGTAGATAT	1860
	ATACTTTTTA AAGCTTGTAG CTAAAGCCTT TATTTAACTG GTTTTGAAAT TTGTGTTTTA	1920
20	CCACCCATAA ATGGTACTAA TGCTTCTGGA ATTGTTACTG TTCCATCTTC ATTTTGGTAA	1980
	TTTTCAACAA TAGCAGCAAA TGTACGTCCA ACTGCTAAAC CACTACCATT TAATGTATGT	2040
	GCTAATTCTG GTTTAGCTGC TTTGTCACGC TTGAAGCGGA TGTTAGCACG ACGCGCTTGG	2100
25	AAATCCGTAC AGTTTGAGCA TGAACATAAT TCTTTATAAT CATTGTAGCT TGGTAACCAA	2160
	ACTTCTAAAT CATATGTTTT GCTTGCACTA AATCCAATAT CACCTGTACA TAAAATAACA	2220
	CGACGGTATG GTAAACCTAA CTCTTCTAGA ATTGCTTCTG CGTTTGTGTG CATTTCTTCT	2280
30	AAAGCATTCC ATGAATCTTC AGGTTGTTC A AACGTACCA TTTCCACTTT ATCGAATTGA	2340
	TGTAAACGAA TTAATCCTCT TGTATCTCTA CCTGCTGATC CTGCTTCACT ACGGAAACAT	2400
	GCAGATTGAC CAGTGAATTT TTCAGGAAGT ACACCTGGTT GAATAATTTT ATTACGGTAG	2460
35	AAATTCGTTA ATGGTACTTC AGCAGTTGGA ATTGTATATA ATCCTTCTTT TTCTACTTTA	2520
	AATAAATCTT CTTCAAATTT AGGTAATTGA CCTGTACCAT ACATTGTATC TGC GTTCACA	2580
	AGCTGTGGTA CCATCATTTT TGTATAACCA TGTTGTGTGT TATGTTTTGT AATCATATAG	2640
40	TTCATTAAAG CACGCTCTAA TTGCGCACCT TCATTGTGTA AATATACAAA ACGCGCACCT	2700
	GAAACTTTTG CTGCACGATC AAAATCAGCC ATTTTCAATT CTTCTACAAT ATCCCAATGT	2760
45	GCTTTGGGTT CAAATGAAAA CTCaGTGGT GTACCCCACT TTTTAACTTC AACGTTATCT	2820
	TCATCAGATT CACCTTGAGG TACATCATCA CTTATTAAAT TTGGAATACG ACAAAGGATA	2880
	CCTGTCATTT TATTATCAAT TTCATTAAAT TGACTATCTT TTTCTTTAAT ATCGTCACCT	2940
50	AATGTGCGCA TTTCAGCAAT CACATCATCA GCATTTTCTT TATTACGTTT TTTAATGCG	3000
	ATTTCTTCGC TTACTTTATT ACGACGTGCT TTCATTTCTT CTGTTGCACT AATTAATTTA	3060

55

TCAATTTTGC TCTTAACTGT GTCAGGCTCA TTTCTGAATA ATCTAATGTC TAACATTAAC 3180
 CTTTCATCCTT TCCCAAATAA TTATCATTTA TTATGGAATG ACGTACGTCT TTATTTTTTA 3240
 5 GAAAATAAAA AAAGACCACA TCCCTACAAG GGACGTGGTC TACGCGTTGC CACCCTATTT 3300
 AACAATTTAA GTTATAAAGA TACACTAAAC CTAAATTGCA CTTCACTAAA ATAACGGTTA 3360
 TCACCGATTG TTCTTTTAAA TTAAGTAGGT AGATTTCATAT ATATGTTGAT TCTTGTTTAC 3420
 10 ACTAACCACA AGCTCTCTGA TATCGAACAC TATATATTAC TTGTCCTACG AACAAATGTCT 3480
 TATTAAGTTA TTTTAAATAT AGCAAACAT ATTTGCTTTT TCAAGTAACG ATTTCAAACA 3540
 TCACTCATGT CGATTAGTG ACATGCAGTC GTTTGATAAA TTGATTGCTT TAAATACTGT 3600
 15 GCAACCGCTT CAATATCTTT ATGAAATTGA CGATCATGTG TAATGGATGG CACGATACTT 3660
 CGAAACTCAT CATACTTGCG ACGTGTTTTT GGTGATAATC CTTCAACACC TTTTAACTCT 3720
 GCTGCTTGTA ATGCAATAAC ACATTCGATT GCCAGCACAC GTCTTGCAAT TTCAATAATT 3780
 TGATAACCAT GTCTAGCAGC TGTAGTTCCC ATAGATACGT GATCTTCTTG GTTCGCAGAT 3840
 GAAGTGATAG AATCAACACT CGCTGGATGC GCTAAAGTTT TATTTTCAGA AACGAGACTT 3900
 25 GCAGCAGCAT ATTGCATAAT CATCGCGCCA CTTTGCAATC CTGGCTCTGG ACTAAGAAAT 3960
 GCTGGTAAAT CACCATTAA TTGAGGATTT ACTAGTCGCT CTAGACGACG TTCCGATACG 4020
 TTTGCTAATT CACTTACACC TAATTTAAGA TGATCTAATG CAAAAGCAAT AGGTTGTCCA 4080
 30 TGGAAAGTTAC CACCTGAAAT AACAAACGTT TCATTTGCTT CCTCAAATAT AAGTGGATTA 4140
 TCATTAGCCG CATTCAATTC AAATTCTAAT TGCTGTTTAA CATAATTGAA TACTTGAAAA 4200
 CTCGCGCCAT GGATTTGTGG TATACAACGC AACGTATATG CATCTTGATC ACGTATTTCT 4260
 35 GATTGTCGCG TCGTTAATGT TGATCCTTCT AACCAATCAC GCATACGCGC TGCCACATTA 4320
 ATCTGTTCTT GAAAATTACG AACTGCGTGC ACATCATGTC GATATGCATC TATAATGCCA 4380
 TTAAGAGACT GATGCGTTAA TGCAGCAATC CATTGAGATT GGTAACCTAA ATCTTCTGCT 4440
 40 TCTATATAAC TAATGACACC TTGAGCTGTC ATAGCTTGCG TACCATTAAAT CAATGCTAAA 4500
 CCTTCTTTAG CCTGAAGGTT CAAAGGTTGT CTATTTAATT CTCTTAATAC ATCGTCACTA 4560
 TCCTTTTCTT CCCCTCTGTA CAATACTTTC CCTTCACCAA TTAATGCTAA TGCTAAATGT 4620
 45 GATAATGGCG CTAAATCTCC TGATGCACCG AGAGAGCCTT GCTGTGGGAT TATCGGTATA 4680
 ATACGTTTCAT TTATAAAAAA TTGTAATGT CTCACCTAAT CTAAAGTGGC ACCTGAATGA 4740
 50 CCTTTTAATA ATGTATTCAA TCGTAAAATC ATCATGACTA ATGCTACTTC TTTTGAAAAT 4800
 GGCTCACCTA GTCCACAGGC ATGTGAGCGT ATCAGATTCA CTTGTAATTC ATTATATTGC 4860

55

	TCCTCATTTT CAATAATACG TTCAACTACC GCTCTACTTT TTTTGACACG TTCTAACGCA	4980
	TCATCAATAA TTTCAATCTT TGATTGTTGT TGTA AAAATG ATTTAATATC CTCAATTGTT	5040
5	AGTGTTCAC CATCTAAATA TAAAGTCATA TATGTTACCC CCTTGTTTAT ATTAAGTAAC	5100
	CCATCCTTCT TGAAGTATAC GTTTTCATTT TTATTGAAAC AATGGTTTTA CGTACATTTA	5160
	TAACCTATTA TCAGAGCACT ATTGTAGTGC GTTAAAGGAT ATTAAGATTG TTGTAAGCAT	5220
10	ATTTAATAAT TTATCTATTG ACGAATTGCA TATACAGGTA TAGTATTTTC TATTGTATTT	5280
	AACGACAAAT AATAATGAAT TCAGAAATTT ATAATACATT TTGTTAAAAG TTACTATATA	5340
	TTTTTAAAAT TGAATAAATT CGGAAAAGGC TTTTACATGG GAGGTTATAT CACTATGGAA	5400
15	ACGTTAAATT CTATTAACAT TCCTAAGCGT AAAGAAGATT CACATAAAGG TGATTATGGC	5460
	AAAATTTTAT TAATTGGTGG ATCTGCTAAC TTAGGTGGTG CCATTATGTT AGCGGCTCGT	5520
20	GCATGTGTAT TTAGCGGTAG TGGTTTAATC ACTGTAGCTA CACATCCAAC AAATCATTCA	5580
	GCATTACATT CTCGTTGCCC AGAAGCGATG GTTATTGATA TTAATGATAC GAAAATGTTG	5640
	ACGAAAATGA TTGAAATGAC TGACAGTATA CTAATTGGTC CAGGTCTTGG CGTTGATTTT	5700
25	AAAGGAAATA ATGCCATTAC ATTCCTACTA CAAAATATAC AACCGCATCA AAATTTAATC	5760
	GTAGACGGCG ATGCGATTAC AATCTTTAGT AAAGTAAAC CGCAATTACC TACATGTCGT	5820
	GTGATCTTTA CACCACACCT CAAAGAATGG GAACGATTAA GTGGTATTCC TATTGAGGAA	5880
30	CAGACATATG AGCGTAATCG TGAAGCAGTT GATCGTTTAG GTGCAACTGT TGTACTTAAA	5940
	AAACATGGTA CTGAAATTTT CTTTAAAGAT GAAGACTTTA AATTGACAAT CGGTAGCCCA	6000
	GCAATGGCGA CTGGTGGTAT GGGCGATACA CTTGCTGGTA TGATTACAAG CTTTGTGCGT	6060
35	CAATTTGATA ACTTAAAAGA AGCGGTTATG AGTGCCACAT ATACACATAG TTTTATTGGC	6120
	GAAAACCTTG CAAAAGATAT GTATGTGGTG CCACCATCAA GACTTATCAA TGAAATACCT	6180
	TACGCAATGA AACAAATTAGA AAGTTAGTCA TTACTAATCA TTGAATATAG TAAAGCATT	6240
40	CTTTCTAGCA TAAAAATAAG ACTCCCCTAC ATATAGGGAA GTCTTATTTT TTATTATTCT	6300
	TCATCTGATG ATTGTTGTAT ATCTTCTTCA ACACGATCCA TGAAATCTTG TCTTACTTCA	6360
	ATACGTCCAT CTTTCATCATT TTCTTCTGAA TCAATCACTT CAGTATGAAT TGCATTTCC	6420
45	GGTGTTCAT CATTTaCAAC CGCTTCACGT TGTGTTCAG TACCATCTTC AGATACAGTT	6480
	GAAGTAGATT GCTCATCTTC ATTCGTTTCA TCTTCTGCAT CTTCTTTTAC TTTAGCAACC	6540
50	GTTGAAACAA ATTGATCATC ACCTAAGCGA ATTAAGCGAA CACCTTGTGC TGCACGACCA	6600
	TTTTGAGAAA TATCTGCAAC ATCTAGTCGA ATAATGACAC CTGCATTAGT AACAAATCATT	6660

55

	GTAGCTGTTT TAATACCTTT ACCACCACGA TTTGATAAGC GATAGTCATT AACTGGCGTA	6780
	CGTTTACCAT AACCATTTTC AGTAACTACT AATACTTCAT CAACACTGTT TGCATGAGCT	6840
5	ACATCAAGCC CTACAACCTC GTCACCTTCA CGAAGTGTA TACCTTTTAC ACCCGTTGCT	6900
	GTACGGCCTA AAGGACGTAA TGTTGATTCA GGAATCGAA TTAATGATGC ATGTGATGTA	6960
	CCAATCAAGA TATCTTCTTG ACCACTTGTT AAGCGAACTG CAATTAAGTC ATCATCTTCT	7020
10	CTGAACGAAA TCGCAATCTT ACCATTCTTA TTTATTCTTG AGAAGTTACT TAATGCTGAA	7080
	CGTTTAACGA CACCACGTTT AGTTGCAAAC ACTAAGAAGT TGTCTTCACT TTCAAGGTCT	7140
	TTAACAGCAA TCATTGTACT AATGACTTCA TCATTTTCAA GTTCAATAGC ATTCACTACA	7200
15	GGAATACCTT TAGACTGTCT TGATAACTCA GGCACCTTCG AACCTTTAAG TTTGTATACA	7260
	CGACCTTTGT TAGTAAAGAA CAATACATGG TCATGTGTAC TTAAAGTTAC CAATTGACTG	7320
20	ACAAAATCTT CTTCCAATGT ATTCATACCT TGAACACCAC GACCACCACG GTTTTGAGCA	7380
	CGATATGTAG ATACCGGCAA ACGTTTAATG TAGTTATTAT GGCTTAGTGT AATTACTATT	7440
	TGTTCTTCTG GAATTAAGTC TTCGTCCTCT AAGTCTTCAA ATCCACCTAA TTGAATTTCT	7500
25	GTACGACGAT CATCACCGAA ACGATCTCTA ATTTCACTCA ATTCATCTCT AACTAACTGT	7560
	AATAACACTT CTTTCATCAGC TAAGATTGCT TCTAATTCAC TAATATAATT TAATAACTCA	7620
	TTATATTAG CTTCAATTTT GTCTCTCTCT AAACCTGTTA GACGTCTTAA ACGCATGTCT	7680
30	AAAATAGCTT GAGCTTGTTT TTCAGAAAGT TTGAAGCGTT GTTGCAAGCT TTCCATTGCA	7740
	ACTTTATCTG TATCTGACTC ACGAATCGTT GAAATAATTT CATCGATATG GTCAAGTGCG	7800
	ATACGTAATC CTTCTAAAAT GTGGGCACGA TCTTTAGCTT TACGTAAgTT GTATTGCGTA	7860
35	CGTCTTCTAA CAACTGTCTT TTGATGCTCT AAATAATGTA CCAACGCTTC TTTTAAATTA	7920
	ATAAGCTTCG GTCTACCATT TACAAGTGCA ATCATATTCA CACCAAATGA TGTTTGAAGA	7980
	GGTGTGTTGTT TGTATAAGTT ATTTAAAATG AACTAGCAT TTGCATCCTT ACGCACATCA	8040
40	ATAACGACAC GCACACCAGT ACGTAACTT GTTTCATCAC GTAAATCAGT GATACCGTCA	8100
	ATTTTCTTGT CACGAACGAG CTCTGCAATT TTTTCAATCA TACGAGCCTT ATTCACTGG	8160
	AAAGGAATTT CAGTGACAAC AATACGTTGA CGTCCGCTC CACGTTCTTC AATAACTGCA	8220
45	CGAGAACGCA TTTGAATTGA ACCACGACCT GTTTCATATG CACGTCTAAT ACCACTCTTA	8280
	CCTAAAATAA GTCCAGCAGT TGGGAAATCA GGACCTTCAA TATCCTCCAT TAACTCAGCA	8340
50	ATTGAAATAT CAGGGTTCTT ACTTAAGCTA AGTACACCAT TGATTAAATC TGTTAAGTTA	8400
	TGTGGTGGA TATTCGTTGC CATACCTACC GCGATACCTG ATGCACCATT GGCTAATAAG	8460

55

	AAATCTATTG TATCTTTATT AATATCACGT AACAGTTCAA GTGTGATTTT AGTCATACGC	8580
	GCTTCAGTAT AACGCATTGC TGCTGCGCCA TCTCCATCCA TTGAACCAAA GTTACCTTGG	8640
5	CCATCAACAA GCGGATAACG ATAAGTAAA TCTTGAGCCA TACGTACCAT TGCTTCATAA	8700
	ATAGATGAGT CACCATGAGG GTGATATTTA CCCATTACGT CACCAACGAT ACGTGCTGAT	8760
	TTTTTATATG ATTTATCCGG TGTCATACCT TGTTTCATTTA ATCCATATAG TATACGACGA	8820
10	TGTACTGGTT TTAAACCGTC ACGAACATCT GGCAATGCAC GAGCAACGAT AACACTCATC	8880
	GCATAATCTA AAAATGATTC ACGCATTTCA CTGGTAATAT TTCGTTTCATT TATTCTTGAT	8940
	TGAGGTAATT CAGCCATCAA GAGTTCCTCC TTCAAAAGTT CAGTTCACAG CGCTTAGAAG	9000
15	TCTAAGTTTG CATAAACTGC ATTATCTTCT ATAAATTGTC TACGGTTTTT TACAACGTCA	9060
	CCCATTAAAC TTTCAAATGT TTGGTCCGCT TCAATCGCAT CTTCAAGTTT TACTTGTAAC	9120
	AGAGCGCGGT GCTCAGGGTT CATTGTTGTT TCCCAAAATT GATCTGCATT CATTTCTCCA	9180
20	AGACCTTTGT ATCGTGCAAT AGACCATTTT GGTGTTGGAT TCAATTCAGA TTAAAGTTTA	9240
	TCAAGTTCCC TATCATTGTA TACATAATAC TTTTGTTTAC CTGTGTGTCAG TTTATACAAC	9300
	GGTGGCTGTG CAATATACAC ATAGCCTGCT TCAATTAACG GTCTCATAAA TCGATAGAAG	9360
25	AATGTTAATA ACAATGTTCT AATATGCGCT CCATCCACAT CGGCATCAGT CATAATGACG	9420
	ATTTTGTGAT ATCTTGCTTT CGCTAGATCA AAGTCGCCAC CGATTCTGT ACCAAATGCT	9480
30	GTGATCATTT GACGAATTC ATTGTTATTC AAAATTCTAT CTAATCGTGC TTTTCAACA	9540
	TTTAATATCT TACCTCGTAA TGGTAAAATC GCCTGCGTTC TAGAGTCACG ACCAGATTTT	9600
	GTAGACCCCC CGGCAGAGTC CCCTTCGACT AAGAAAATCT CACATTCTTC AGGACTTTTA	9660
35	CTAGAGCAAT CGGCTAATTT ACCTGGAAGG CTTGCTACAT CTAACGCTGA TTTACGACGT	9720
	GTTACTTCAC GCGCTTTTTT CGCAGCAACA CGTGACAGTG CCGCCATAAT ACCTTTTTCA	9780
	ACCACTGTAC GTGCGACTTG TGGATTTTCA TATAAAAATC GTTCAAAGTG CTCTGAGAAT	9840
40	AATTTATCTA CAACTTGACG CACTTCAGAA TTACCTAATT TTGTCTTCGT TTGACCTTCG	9900
	AATTGAGGAT CACCATGTTT GATAGATATA APTGCTGTCA TACCTTCACG TGTATCTTCA	9960
	CCAGAAAGTC TATCTTTTTT TTCTTTCATA ATCTTGCTAC TTAAACCATA ACTATTTAAG	10020
45	ACACGCGTTA ATGCACGTTT GAATCCGTCT TCATGCGTAC CACCTTCATA CGTATGAATG	10080
	TTATTTGCGT AAGTTAAAAG ATTTGTGGCA TATCCTGAGT TATATTGAAT CGCAATTCTT	10140
	ACTTCAATAT CATCTTTAGA TTGATGAATA TAAATTGGCT CATCATGAAT AGGTTCTTTA	10200
50	TTTTCGTTCA ATAAGTCAAC GTACGATTTA ATACCGCCCT CATAGTGATA GGAGTCTTCT	10260

55

GCAAGCTCTC TAATACGCTG CTGTAATGTT TCATAGTTGT ATACAGTTGT CTCTGTGAAG 10380
 ATTTCTCCAT CTGCTTTAAA ACGAATGaCa GTACCTGTCT TAtCAGTnGT GCCAACTTCT 10440
 5 TTTAAGTCAA ATTGAGGTAC ACCTTTTTTA TATGCTTGAT GATATATAGT CTCATTTCTG 10500
 TGTACATATA CTTCTAAGTC TTGTGACAAT GCGTTTACAA CTGATGAACC AACACCATGT 10560
 AAACCACCAG ATACTTTGTA TCCGCCACCG CCAAATTTAC CACCAGCATG TAAAACAGTT 10620
 10 AAAATAACTT CGACAGCTGG ACGTCCCATT TTTTCTTGAA TATCAACTGG GATACCACGT 10680
 CCGTTATCCG TTACTTTAAT CCAGTTATCT TTTTCAATAA CAACTTCAAT TTGATTTGCA 10740
 TAACCAGCTA ATGCTTCATC GATACTATTA TCGACAATTT CCCACACTAA ATGGTGCAAA 10800
 15 CCTCTCTCTG AAGTCGATCC TATATACATA CCTGGTCTTT TACGTACTGC TTCTAAACCT 10860
 TCTAATACTT GTATTTGCCC AGCACCATAA TTATCCGTGT TGTTTACATC TGACAATGCA 10920
 20 GTCACCATCG CTTTCTGTTA CTTTATAATT TCACCTTGAT TAATACGATA CAATTTAGCG 10980
 TTATTCATGA TTTCATGATC AATACCATCT ACAGATGTCG TAGTGACAAA TGTTTGTACT 11040
 TTATGCTGAA TCGTACTTAA TAAATGCGTT TGACGCGAAT CATCTAATTC ACTGAGTACA 11100
 25 TCGTCTAATA ATAAGATGGG ATATCCCCA ACTTCGATAT TCATTAACTC AATTTTCAGCT 11160
 AATTTAATGG ACAAAGCCGT TGTACGTTGC TGTCCTTGAG AACCATATGT TTGAGCATCC 11220
 ATGCCATTCA CATCAAACT TATATCATCT CGATGTGGTC CGAATAAGCT AATGCCTCGT 11280
 30 TCTTTTTCTC TTTGCATATT ATCGCTAAGA ATAGACATAA TTTCTTCAAG TCGTGCCGCT 11340
 TCATTTTGAG CATAATCAAA TTTAAGACTA GGTAATAAT TCAGCGACAA CGCTTCTTTA 11400
 TCATTTGTGA TACCAGCATG AATCGTTTGA GCTAACGACT CTAGCTCTTG AATAAAATGT 11460
 35 GCACGTTTAT CAGTTACTTT CATTGCATAT TCAGCAAAT GCTGATTTAA TACTTCCAAC 11520
 ATTGTTAAGT CCTTTTTTTG GCCTAATTGT AACTGCTTTA AGTAATTATT CTTTTGCTTT 11580
 AAAATACGTT GGTATTGAGC TAAATCATTT AAGTAAACAG CAGAAATTTG GCCAACTCC 11640
 40 ATATCTATAA AGCGTCGTCT TATTtGrGGr GAGCCTTTTA CAATATTCAA ATCTTCTGGC 11700
 GCAAATAGAA CCACATTGAG GTGTCCAATA TATTGAGTTA GACGACTTTG CTCTAAGTGr 11760
 ATTCACTTTG GACTTGTTTA CCTTTnTTAG TTATAAACAT TGTTAATGGG CATCGTGCCG 11820
 45 TGT 11823

(2) INFORMATION FOR SEQ ID NO: 137:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 692 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

5 ATAATTATTA ACATGGTGTG TTTAGAAGTT ATCCACGGCT GTTATTTTGT TGTATAACTT 60
 AAAAAATTTAA GAAAGATGGA GTAAATTTAT GTCGGAAGAA GAAATTTGGG AAAAAAGTGCT 120
 TGAAATTGCT CAAGAAAAAT TATCAGCTGT AAGTTACTCA ACTTTCCTAA AAGATACTGA 180
 10 GCTTTACACG ATTAAAGATG GTGAAGCTAT CGTATTATCG AGTATTCCTT TTAATGCAAA 240
 TTGGTTAAAT CAACAATATG CTGAAATTAT CCAAGCAATC TTATTTGATG TTGTAGGCTA 300
 TGAAGTTAAA CCTCACTTTA TTACTACTGA AGAATTAGCA AATTATAGTA ATAATGAAAC 360
 15 TGCTACTCCA AAAGAAACAA CAAAACCTTC TACTGAAACA ACTGAGGATA ATCATGTGCT 420
 TGGTAGAGAG CAATTCAATG CCCATAACAC ATTGACACT TTTGTAATCG GACCCGGTAA 480
 CCGCTTTCCA CATGCAGCGA GTTTAGCTGT GGCCGAAGCA CCAGCCAAAG CGTACAATCC 540
 20 mTTATTTATC TATGGAGGTG TTGGTTTAGG aAAAACCCAT TTAATGCATG CCATTGGTCA 600
 TCATGTTTTA GATAATAATC CAGATGCCAA AGTGATTTAC ACATCAAGTG AAAAATTCAC 660
 AAATGAATTT ATTAAATCAA TTCGTGATAA na 692

25 (2) INFORMATION FOR SEQ ID NO: 138:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7900 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

35 ATACTGTAGC GCAAATTTCA CAATGGCATG TTATAGAAGA TTTAGTTACG AATGAATTAG 60
 GTATTAGTAT TTTACCAACA TCAATTTAG AGCAACTAAA TGGAGATGTG AAGCTGTACG 120
 40 CATTGAAGAT GCTCATGTAC ATTGGAATT AGGTGTTGTT TGGAAGAAGG ATAAACAATT 180
 AAGTCATGCC ACAACGAAAT GGATAGAATT TTTGAAAGAC CGTTTAGGCT AACATATTAA 240
 TAAAGCACTC ATTATTTAAG GCGCATCATT ACGTGGGTCA TTGAAATAAT GAGTGTTTTT 300
 45 TTGTGAAAAT GAAGTGAAAT TTAGAGAGCG TTTCCATAGA AAATAGTAAT ACAAATAA 360
 AAAAAAGAGT ATTTTATAT TGTGTACGCC ATCTTTATAA TAGTTATTGT ACAAATTTAG 420
 ACATATTTAG AAAGGGATGG CGCCATGCAC AAAGTCCAAT TAATAATCAA ACTACTACTA 480
 50 CAACTAGGAA TCATCATTGT GATTACTTAT ATTGGCACAG AAATTCAAAA GATTTTTCAT 540

55

	ATTGTACCGC TAACTTGGGT AGAAGACGGT GCAAACCTTTT TATTAAAGAC GATGGTCTTT	660
	TTCTTCATAC CGTCAGTTGT AGGtATTATG GaTgTgCTTC CGAAATTACG CTAAATTATA	720
5	TACTCTTTTT CGCAGTCATT ATCATAGGAA CATGTATCGT TGCATTATCT TCAGGTATA	780
	TTGCTGAAAA AATGTCyGtT AAACWTAAAC ATCGTAAAGG TGTAGACGct TATGAATGAT	840
	TACGTGCAAG CCTTATTAAT GATTTTGTG ACTGTGCTTT TATATTATTT CGCTAAAAGG	900
10	TTACAACAAA AATATCCGAA CCCATTTTGTG AATCCAGCAT TAATTGCATC TTTAGGAATT	960
	ATTTTTGTCT TACTTATCTT TGGAATTAGT TATAACGGGT ATATGAAAGG TGGCAGTTGG	1020
	ATCAACCATA TTTTAAACGC AACGGTCGTA TGTTTAGCGT ACCCACTTTA TAAAAATAGA	1080
15	GAGAAAATTA AAGACAATGT CTCTATCATT TTTGCAAGTG TATTAAcTGG CGTCATGCTG	1140
	AATTTTCATGT TAGTGTCTT AACACTTAAA GCATTTGGCT ATTCTAAAGA CGTCATTGTA	1200
	ACGTTATTGC CCCGATCTAT AACAGCCGCA GTAGGTATCG AAGTGTCA CA TGAAC TAGGT	1260
20	GGTACAGATA CGATGACCGT ACTTTTATT ATCACAACGG GTTTAATCGG TAGTATTTTA	1320
	GGTTCGATGT TATTAAGATT TGGAAGATTT GAATCTTCTA TCGCCAAAGG ATTAACGTAT	1380
	GGGAATGCGT CACATGCATT TGGCACAGCT AAAGCACTAG AAATGGATAT TGAATCCGGT	1440
25	GCATTTAGTT CAATTGGGAT GATTTTAACT GCAGTTATTA GTTCAGTGTT AATACCTGTT	1500
	CTAATTTTAT TATTCTATTA ATTTAGATAT TTAAATGAT AGACAGAAAG GGAGGCTATT	1560
30	AGTAATAATG GCAAAAATAA AAGCAAATGA AGCATTAGTT AAAGCATTAC AAGCaTGGGA	1620
	TATAGATCAC TTGTATGGTA TTCCAGGAGA CTCAATCGAC GCATAGTCGA TAGTTTACGT	1680
	ACAGTGAGAG ATCAATTTAA ATTTATCAT GTACGTCATG AAGAAGTAGC AAGCTTAGCG	1740
35	GCTGCTGGTT ACACAAAATT AACTGGTAAA ATCGGTGTGG CATTAGTAT CGGTGGCCCT	1800
	GGTTTAATTC ATTTATTAAA TGGTATGTAT GATGCCAAAA TGGATAATGT ACCGCAATTA	1860
	ATATTATCTG GACAAACGAA TAGTACAGCA CTTGGAACGA AAGCATTCCA AGAAACAAAT	1920
40	TTACAAAAAT TATGTGAAGA TGAGCCGTT TATAATCACC AAATTGAAAA AGGTGACAAT	1980
	GTGTTTGAAA TCGTTAACGA AGCAATTCGT ACGGCATATG AACAAAAAGG TGTAGCTGTT	2040
	GTTATTTGTC CTAACGACTT ATTAACGTAA AAAATTAAAG ATACAACGAA TAAACCAGTA	2100
45	GATACATCAA GACCAACAGT AGTATCACCA AAATATAAAG ACATCAAAAA AGCGGTAAA	2160
	CTAATTAATA AAAGTAAAA GCCTGTCATG TTAATTGGTG TAGGTGCGAA ACATGCGAAA	2220
	GATGAGCTAC GTGAATTTAT TGAAATGGCT AAAATTCCTG TCATTCAATC ATTACCAGCT	2280
50	AAAACAATCT TGCCGGATGA TCATCCATAT AGTATCGGtA ACTTAGGTAA AATCGGTACC	2340

EP 0 786 519 A2

	CCATATGTGG ATTACTTACC TAAGAAAAAT ATTAAAGCCA TTCAAATTGA CACAAATCCT	2460
	AAAAATATCG GACATCGTTT CAATATTAAT GTAGGAATTG TTGGAGATAG TAAAATTGCG	2520
5	TTGCATCAGT TAACTGAAAA TATTAAACAT GTTGCTGAAA GACCATTCTT AAACAAAACG	2580
	TTAGAACGTA AAGCGGTTTG GGATAAATGG ATGGAACAAG ATAAAAATAA TAATAGTAAA	2640
	CCATTACGTC CAGAACGATT AATGGCATCA ATCAATAAAT TTATTAAAGA TGATGCAGTG	2700
10	ATTTACAGCAG ATGTAGGTAC AGCAACAGTT TGGTCAACTC GATACTTAAA CCTTGGTGTA	2760
	AATAACAAGT TCATCATTTT AAGTTGGTTA GGTACAATGG GTTGCGGTCT TCCAGGTGCA	2820
	ATTGCATCAA AAATTGCATA TCCAAATAGA CAAGCCATCG CAATTGCTGG TGACGGTGCA	2880
15	TTCCAAATGG TAATGCAAGA CTTGCTACA GCAGTACAAT ATGATTTACC TTAACTGTA	2940
	TTTGTACTTA ATAACAAACA GTTAGCATTT ATTAAATATG AACACAAGC AGCTGGTGAA	3000
	TTAGAATATG CAGTTGATTT TTCTGATATG GATCATGCAA AATTGCTGA GGCAGCAGGT	3060
20	GGTAAAGGTT ATACAATTAA GAGTGCTAGC GAAGTAGATG CTATAGTCGA AGAGGCATTA	3120
	GCACAAGATG TACCAACGAT TGTAGATGTA TATGTTGATC CTAATGCTGC GCCATTACCA	3180
	GGTAAATTG TAAATGAAGA AGCGCTTGGT TATGGTAAGT GGCATTTAG ATCAATTACT	3240
25	GAAGATAAAC ATTTAGATTT AGATCAAAT CCACCAATTT CAGTGGCAGC AAAACGTTTC	3300
	TTATAACTGA TTAAAGGTT ATCACAATTG AATTGAACTA TAAAACGGT AATTTCTATT	3360
	TCAACAAAAT GGAATTGCC GTTTTGTTTA TTTATCACAA ATGATCGTAC TGAATTGATG	3420
30	ATAAAATTGT GAAAAAGTTG TTGAAAACGC TTTTACAAAT ATGTATAATA GCTATGAATT	3480
	AGATATCACT TGCGTGTTAC TGGTAATGCA GGCATGAGCA AACAACCGCA CTATGAGAAT	3540
35	AGTCTTGTTT GTTCATGCCT GCTTTTTTTG TACATGGAAG CGGAAATTGA GATAGGGGAT	3600
	GTTTETATGT TTAAGAAAT GTTTGGACAA TTGCAACGTA TCGGTAAAGC ATTAATGTTA	3660
	CCTGTTGCGA TTTTACCAGC AGCTGGTATT TTATTAGCGT TTGGTAACGC AATGCACAAC	3720
40	GAACAATTAG TAGAAATTGC ACCATGGTTA AAAACGATA TCATTGTAAT GATTCGTCG	3780
	GTCATGGAAG CAGCAGGACA AGTTGTATTT GATAACTTGC CATTATTATT TGCAGTTGGT	3840
	ACAGCACTTG GATTAGCAGG AGGAGACGGT GTTGCAGCAT TAGCAGCGCT AGTAGGTAC	3900
45	TTAATTATGA ATGCAACAAT GGGGAAAGTG TTGCACATTA CAATTGATGA CATTTTCTCA	3960
	TATGCCAAAG GGGCAAAGA ATTAAGTCAA GCAGCGAAAG AACCAGCACA TGCTTTAGTA	4020
	TTAGGTATTC CAACGTTACA AACGGGTGTG TTTGGTGGTA TTATCATGGG TGCTTTAGCC	4080
50	GCATGGTGTT ACAACAAATT TTATAATATT ACACTACCAC CATTTTTAGG ATTCCTTGCA	4140

55

	AGCTTTGCGT GGCCACCAAT TCAAGATGGA TTAAATAGTT TATCGAATTT CTTATTAAAT	4260
	AAAAATTAA CATTAAACAAC GTTATATTC GGTATTATTG AACGCTCATT AATTCCATTT	4320
5	GGTTTACATC ATATTTTCTA TTCACCGTTC TGGTTTGAAT TCGGAAGTTA TACAAATCAC	4380
	GCAGGTGAAT TGGTTCGTGG TGACCAACGT ATTTGGATGG CACAATTGAA AGATGGCGTA	4440
	CCATTTACTG CTGGTGCATT TACTACTGGT AAATATCCAT TTATGATGTT TGGTTTACCA	4500
10	GCGGCGGCAT TTGCTATTTA TAAAAATGCA CGACCAGAAC GTAAAAAAGT CGTGGGTGGT	4560
	TTAATGTTAT CAGCAGGATT AACTGCATTT TTAAGTGGTA TCACTGAGCC ATTAGAATTT	4620
	TCATTCTTAT TTGTAGCACC AGTACTTTAT GGAATTCACG TATTATTAGC TGGTACATCA	4680
15	TTCTTAGTAA TGCATTTATT AGGCGTTAAA ATTGGTATGA CATTCTCAGG TGGTTTCATA	4740
	GATTATATTT TATATGGTTT ATTAACTGG GATCGTTCAC ACGCATTATT AGTTATTCCA	4800
	GTCGGTATTG TATATGCTAT CGTGTATTAC TTCTTATTCG ACTTTGCAAT TCGTAAGTTT	4860
20	AAATTGAAAA CACCAGGTCG TGAAGATGAA GAACTGAAA TTCGTAACTC TAGTGTGCGA	4920
	AAATTACCAT TTGATGTCTT AGATGCAATG GGTGAAAAAG AAAACATTAA ACATTTAGAT	4980
	GCATGTATTA CACGTCTACG CGTAGAAGTG GTTGATAAAT CAAAAGTAGA TGTAGCAGGT	5040
25	ATTAAAGCTT TAGGCGCATC AGGTGTATTA GAAGTTGGAA ACAATATGCA AGCTATCTTT	5100
	GGTCCAAAAT CAGATCAAAT TAAACATGAT ATGGCCAAGA TTATGAGTGG TGAAATTACG	5160
30	AAACCAAGTG AAACGACAGT GACTGAAGAA ATGTCAGATG AACCAGTTCA CGTAGAAGCA	5220
	CTTGGAACAA CAGACATCTA TGCACCAGGT ATCGGTCAAA TCATTCCATT ATCAGAAGTA	5280
	CCTGATCAAG TATTCGCTGG TAAATGATG GGTGATGGTG TTGGCTTTAT CCCTGAAAAA	5340
35	GGTGAAATTG TAGCACC GTT TGATGGTACA GTGAAAACAA TCTTCCCTAC GAAACATGCG	5400
	ATAGGATTAG AATCTGAAAG TGGCGTCGAA GTACTTATTC ATATTGGTAT CGATACAGTG	5460
	AAACTGAATG GTGAAGGATT CGAAAGTCTG ATTAACGTTG ATGAAAAAGT AACACAAGGT	5520
40	CAACCATTAA TGAAAGTGAA TTTAGCATAC TTGAAAGCAC ACGCACCAAG CATCGTTACA	5580
	CCAATGATTA TTACAAATCT TGAAAATAAA GAACTTGTCA TTGAAGATGT ACAAGATGCT	5640
	GATCCAGGTA AGCTAATTAT GACAGTCAAA TAATGATTAA AAATGAAACA GCATATCAAA	5700
45	TGAATGAACT TTTAGTCATT CGTAGTGCCT ATGCGAAGTA GCGAGTTGAA AGAGAATACG	5760
	TTACAAAAGG CAGTAGCTTA AAATGAAGCT ACTGCCTTTT TAGTGCGCAA TGATGTATAG	5820
	CAGGTGTGTT GATGTAATA AGTTAAATAT TAGTGTTAGA TATAGAAAAC ATTGCTTATG	5880
50	TTTTTGTCAC ATTTTAGAAA AATGCATCTT CGCGACTAGC CAAATTAATA GTCTCATTGA	5940

55

	AATAAATTAA CATGATTTTA AATCTATTTG TAAGATAAGG AGATTTGTCA TTATGACAAC	6060
	AGAAGGTCTA TTAGTTGCAG AGAAAGAAAT CGAAGTGAAT GGTACGACA TTGATGCGAT	6120
5	GGGTGTCGTT AGTAATATCG TTTATATTAG ATGGTTCGAA GATTTGAGAA CAGCGTTTAT	6180
	TAATCAGCAC ATGAATTACT CAACAATGAT CAATCAAGGC ATTTACCTA TACTTATGAA	6240
	AACGGAAGCA GAGTATAAAG TACCTGTCAC AATACATGAC AAACCAGTAG GTCGTATTTA	6300
10	CTTAGTTAAA GCAAGCAAGA TGAAATGGGT GTTTCAGTTT GAAATTGTGT CCGCACATGG	6360
	CGTGCAATTGT ATTGGTACAC AGACAGGCGG TTTTACAGA TTGAGTGATA AGAAGATAAC	6420
	CTCTGTGCCA CAAGTGTTTC AAGACATTTT AGCAACAAA TAATGACTTC ATTTTAAAT	6480
15	ATAAAAAGTA AGAAGGTGTT CGAAATGGTT AAGCAATTAA ATAGTGTGCA AGCATTCCGT	6540
	GAATTTATTC ATCAATATCC GTTAGCAGTT GTACATGTCA TCGCGATCA GTGTAGCGTG	6600
	TGTCATGCCG TTTTACCACA AATTGAAGAC TTGATGCAAT CATATCCCA TGTGCCATTA	6660
20	GCTGTGATTA ATCAAAGTCA GGTGGAAGCT ATTGCTGGAG AATTAAATAT TTCTACTGTA	6720
	CCTGTGGATT TAATTTTTAT GAATGGAAAA GAAATGCATC GTCAAGGGCG TTTTATCGAT	6780
	ATGCAACGTT TTGAACATCA TCTTAAGCAA ATGAATGATA GTGTAAATAA CGATGTCGAT	6840
25	GAGCATTAAAT ATCGCAAATG ATTAGCATTG CTAAGATTAT GTAGACATCA TAACTTATTT	6900
	CCCAGTAAAT ATTGGTAGTA ATTAGAATCA GCATGGTACA GTAGAACTAT AGTAGAAATC	6960
	ATCAAAGAGG AGTGACGACA AATGCGTAAA AAATGGTCTA CACTTGCGTT TGGATTTTTA	7020
30	GTTGCAGCAT ACGCACATAT TAGAATTTAA GAAAAACGCA GTGTGAAAAG TTATATGTTA	7080
	GAACAAGGTA TACGATTATC TAGAGCTAAG CGTCGTTTTA TGTATAAAGA AGAAGCGATG	7140
35	AAAGCATTAG AAAAAATGGC GCCACAGACA GCAGGCGAAT ATGAGGGAAC CAATTATCAG	7200
	TTTAAGATGC CAGTAAAAGT GGATAAGCAC TTCGGTTCAA CCGTTTATAC CGTTAACGAT	7260
	AAACAAGATA AGCATCAACG CGTTGTATTA TATGCACATG GAGGCGCATG GTTCCAAGAC	7320
40	CCACTCAAAA TTCATTTCGA ATTTATTGAT GAACTTGCGAG AAACACTCAA TGCTAAAGTC	7380
	ATCATGCCAG TATATCCGAA GATTCCGCAT CAAGATTATC AAGCGACGTA TGTGCTTTTT	7440
	GAAAAGTTGT ACCATGATTT ATTGAATCAA GTAGCAGATT CTAAACAAAT CGTTGTAATG	7500
45	GGTGACTCTG CGGGCGGTCA AATTGCTTTA TCATTGCTC AATTGTTAAA AGAAAAACAT	7560
	ATTGTGCAAC CAGGACATAT TGTATTAATT TCACCAGTTT TAGATGCAAC GATGCAGCAT	7620
	CCTGAAATTC CTGACTACTT AAAGAAAGAC CCAATGGTAG GTGTGGATGG CaGTGTGTTC	7680
50	TTAGCTGAAC AATGGGCAGG GGACACACCT TTAGATAACT ACAAAGTATC ACCAATTAAT	7740

55

CCAGATGCTT TGAAGTTATC GCAATTGTTG AGTGCAGAAAG GTATCGAACA TGACTTTATA 7860
 CCTGGATATT ACCAATTCCA TATTTATCCA GTATTTCCGA 7900

5 (2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1984 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

15 GTCTAAATAA AAAAAATTAT CATTGATTaC TGAAGTGGCA TTTCGAAGTA ATGCTTCAAT 60
 ATCATTGCGAA TATTTCTTCA ATTTATGATT GTGAAATAAT TCTTGCATCA AAAATGGTCT 120
 TTGGTCACAT GAATGTGCAT CTGAAGCTAC AAAATGAGCC AAATTACATT CTATAAATTG 180
 20 TAATGATAAC TTTTGAATGT TTTTACCAAA TCCACCAACT AAAGAACTCG ATGTTAATTG 240
 ACTCAGTGCC CCATTTGCAA CCAATTCATA TAATATTTCC GGATTTTTTG CGATACTTCT 300
 ATTTCTTTCA GGATGTGCAA TGATTGGTAT GTAACCTCTC GATTGTATTT CAAAAACAA 360
 25 TTGTTTTGTA TAATGTGGTA CTTGCCCCGT TGGAAATTCA ATTAATAAAT ATTTGGAACG 420
 ATTAATACCT TGAATACTAC CATTATCTAA GCCTTTCAGA ATCGAATCTG TAATTCTAAT 480
 30 TTCTTGCCCG GGAAATAATT TAATATCCAA TGCTTGAAGT TCTGGATGCG TTCTTAACTC 540
 CGCCAATTTT ACAAGCACTT GTTGAAATGT ATTATCATAT CTCGGATGCA AATGATGAGG 600
 TGTCGCTACA AACTTGTGTA CACCTTCATC CTTAGCTTGC TTTAATAGTG CAATACTCTT 660
 35 TTCAATTGTT TTAGGACCAT CATCTATATC AACTAATATA TGGTTATGAA TATCAATCAT 720
 GATTCATCAG TCCCATATAA TGCATAGTAA CTAGCACTTT TATCTTTAGG CATTCTATTT 780
 AAGACTACAC CTAATAATTT AGCACCTGTT GCTTCAATAA GTTCTTTTCC TTTTTTAACT 840
 40 TCATCTCTAT TATTATTTTC CGAATTAACT ACGTAGACAA CATTGCCGGT AAACCTTGAA 900
 AATAATTGCG CATCTGTAAC TGTGTTCACT GGTGGCGTAT CGATAATTAC AAAGTTATAA 960
 TTCATCAATA ATGTGTCATA CAAATTGCA AATGCCCTTG ATGTAATTAA CTCTGACGGA 1020
 45 TTCGGTGGGA TTGGCCCAGA CGTCAAGACG TCTAAATCTT GAATTTGAGT TGAGATAATA 1080
 CTGTCTTGAT AAGTTGACCA ATTTAGCAAT AAACCTGATA GGCCTTCATT GTTTGGCAAA 1140
 TTAATAATAT AATGCTGCGT AGGTTTACGC ATATCCCCGT CTACGATTAG TGTTTTATAA 1200
 50 CCTGCTTGCG CATATGCAAC TGCTAAATTT GCTGCAATTG TAGACTTACC TGCGCCTGGT 1260

55

5 GATCTTATGC CTCGAAATTT CTCGCTAATA GGTGACTTTG GTTGTTTCATG GACAATTAAA 1380
 CTTGATGTAC TTCyTCGTGT ATTCGTCATG GTAATTCCTC GTAAATTAAA ATTTTTGTAT 1440
 TGAACCTAAA ATAGGTAATC CTAGTTGCGA TTCAACATCT TCTTCTGTCT TAATACGCTT 1500
 ATCTAATAAT TCTTTTAAGA AAATAATCAA TATTGCTAAA ACAATACCAA CAATAATGCT 1560
 GATAACTAAG TTGACAGATA CTATTGGAGA TACTTTTACA GCATTATCAT GTGCTGAGGA 1620
 10 AAGTATCGTA ACATTATCAA CACTCATAAT TTTAGGCATG TCATGAGCAA AAACCTTTAGA 1680
 TATTTTATTA ACAATTTTGT CAGATTGAGA TTTATTCCCA GTGGTAACTG ATACAGTAAT 1740
 AATTTGAGAG TTTGTTTGAT TGGTTACTTT TAAAAATGAA TTCAACTCAG CTGTTGAATA 1800
 15 CTGACCATCA AATTCTCTAG ATACTTTATC TAGAATTCTA GGACTTTTGA TAATTTCCGT 1860
 ATATGTATTA ACAGACTGCA AACTACTTTG AACATTTTGG AAAGCTAAAT CACTTGAGGA 1920
 CTTTTTCATG TTCACTAATA TTTGAGTAGA AGCAGTATAT TTGTCAGGCA TAACAAAAAA 1980
 20 GGTT 1984

(2) INFORMATION FOR SEQ ID NO: 140:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6272 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

CAAATCCCTT GGTGATGAtA AAtGtATTGC TGTGTAGCCA AATAATCTTC GTATATATGA 60
 35 CTGACGTTCA ACAACAGCTT GCAATCGTTT CGTTGGTACA GTTACTTTCT TCTTGTTAAA 120
 GAGACCATAT TCAATTTTAA GTTGCTCATT TTCAAGCATC ACCGAAAAGC CATAAAATCT 180
 TATCATTGTT ATAATCGTTC CAATAATATA TGCCACTATT AATACTAGTA AAATGATGAT 240
 40 TAATACTGAA ATACTTACAA TTTGAACCCA TTGACTAATT TCATGATTTA GCTTCGACCA 300
 TGGGATCAAC TCTCTTACAG CCCCGTAAAT CGGTACTAAA GCTGCTAACG TTACACCAAT 360
 GGCGCCACTG GTCATTGCCA TAAATAGTGA TTCTTTAAAA TTCATCTGAT ATATAGGAAT 420
 45 GCGTTTATTT TTCTGATTAA GCATACTATC AGTGTCTGTC ACTTCATCTA AGCGACCTTC 480
 TGCGATGTCT TCCACATTAC CTTCAATGTC ATGATTACAG TTGTCATTCT TCTCAGCACT 540
 AGACTTTTGC GCCACTTCTG TCTTCAACTC TGTTTGCAAT TGATCAATAT ATCGTTCAAG 600
 50 ATATTACCT TGTTTTTTCG AAATAACACT TAAGACAATA CCATCACTTG GTGTTTGTAT 660

55

	AAATACGTTTT ATATTTAATT CTTTACGCTT TTTATTAAAA ATACCTGTTG TTAAATGAA	780
	ATAATTATCC tCAATCCAAT ATCGCGTGT CATAATTCCG ACAATTTGAG AAATGTATGA	840
5	TATTAAAAAG AATACAAATA CAATACCTAT CCATAAATAT GATTCCGGGAT TCGTATAATC	900
	AAAATCTTTC AATTGAAAGA TAATGAAAAT AAAAAAGACG ACTATGTTTT GTTTGATAGC	960
10	ATTGATTATG CCATTAAAAT ATGAAATCGG ATGTAATTTT TGAGGTTTCAG ACATCACTTT	1020
	CAACCCCTCT CAAATTCGAC ATAGTTCTCT CTTTCGATTAT TTTAACATCG TCATGAGACA	1080
	TCATCGGTAA ATAAATAGTA TGACCTGCAG TCATAAATCC AACTTTATAC AAATTAAGCA	1140
15	CTTTACTAAT TGGATTAGAT TTAATCGACA AGTATTGTAA ACGTTCAATT CGACTCGTTT	1200
	CTTCTTTATA TATAAAAAAT GATGTACGAT ATTGTACACT TAGTTGATCA ACTTTATAAA	1260
	AGCGACAATG ATATTGCCAT AAAGGCTTAA TAAATAATTT TAATGTACTC AGAGCACCTA	1320
20	AAACCAACAA AATATAAAGT AAGTAATGTG GCCATTCAAA TCTTAACCAT ATAAAATAAA	1380
	AAATGACATA CACAGCTACA CTCAATATAA ATTCTAAGCC ATTCGTAATG TAGTAATACA	1440
	ACAATGCTGA CTTAGGACTC TTAGTCAACT TAGTATAATC TGACATATAC CCCTCTCCCC	1500
25	AAATAAAAAA TTATACGGAT TTATAATCTA TTTTCATTTA TTTTATATG ATGATAATTA	1560
	TAGCATATGG AATATTTTAT GCTAATTTAT TCTTCCTAAA GGTACATCTA AAAATTTAAT	1620
30	TAAGCAGAAA GTGCTTGAAT TGCTAAAAAG ACACCATGTT ATAATTTTAT CAACATGATG	1680
	CCTTTCATCT ATAATCAATC TTTTCATCTTA TCAAGAGCGA TATTTAGTTC AAGCACATTC	1740
	ACATAATCAT TTGTTAACAC ACCACGCTGC TTACGATGTT GAATCAAGTC GGCCACTCTT	1800
35	GAAGTAGATA CATGACGAGC ATCAGCAATA CGAGGTGCTT GCTTCAATGC ATTTTCGACC	1860
	GTAATATGCG GATCTAAGCC CGACCCAGAA CTGTGTGCAG CATCTATTGT TACATTTGAA	1920
	TTCCCAAATT TAACATGATG TTTTCATGCGT GCTATTAATT CGGTGTTTCC ATTCGATTCA	1980
40	TTACTTCCAC CTGAAGATAC GCCGTTTTTA TATAATTTTT CAGGATTCAT ATTATAATCA	2040
	ACTGCACTCG GTCTCCCGTG AAAATATCGT GTCTCTGTCC AGTGCTGTCC AATCAATTTT	2100
	GATCCAACATA TACGATTGTC ATACGTAATT AAACCTGCCAT TTGCTTGTTG ATAAAAAAT	2160
45	ATTTGACCAA TTAACGTGAT AGCTAACGGG AATAAAAATC CACATAATAC CATAGTTATT	2220
	ATCGTTAAAC AAATACTATT TCTTATCGTA TTCATGGTAC AGGCTCCTTC CTCTTTACAC	2280
50	AAAAAATTGT ACAATCATAT CTATTAATTT AATGCCTAAA AACGGGACGA TTAATCCACC	2340
	TAATCCATAA ATCAACATAT TATTTATAAA GATTCTATCA ATGCTGTAAC CCTTTACTTT	2400
	TACACCTTTC ATGGCAATTG GAATTAAGGC AACAATGATT AATGCATTGA ATATCAAAGC	2460

55

	AATTGTTGAC ATCATTAGTG CAGGTAAAAT TGCAAAGTAT TTTGCTACGT CATTAGCCAA	2580
	ACTAAATGTC GTTAATGCAC CTCTCGTCAT TAATAATTGT TTGCCTATTT TTACAACCTC	2640
5	TATTAACTTT GTAGGATTCTG AATCTAAATC AATTAGATTA GCTGCCTCTT TAGCACTAAT	2700
	TGTCCCTGAG TTCATAGCTA ATCCTATATT CGCTTtGTGc tAGCGCAGGT GCATCATTTG	2760
10	TACCATCTCC TGTCATCGCA ACAATATGGC CTTTCGCTTG TTCATCTTTG ATGACTTTAA	2820
	TTTTATCTTC GGGTTTACAC TCTGCAACAA ATCTATCAAC CCCGGCTTCT TTTGCAATTG	2880
	TAGCTGCTGT TAAAGCATTa TCACCTGTAC ACATAACTGT TTCAATCCCC ATTTTTCTCA	2940
15	ATTCAGTAAA TCGTTCTACA AGACCATCTT TAATCACATC TTTTAAATAA ATCACGCCAA	3000
	GCATGACATT GTTTTCAATG ACTATTAAatG GnGTGCCACC TTTACTCGAT ACATCCATAC	3060
	AGAGAGACTC AATATTAAGA GGAATATTGC CTTGTTGTTG TTTGACAAGA TTTATCATAC	3120
20	TATTAGGTGC ACCTTTGAAT ACCGATATTT CATTTGTAAT GATTCCGCTC ATTCTAGTTT	3180
	CAGCTGTAAA AGGCTTATAT GTGCCATCAA TGTCTTTAGG CAGCTCATTT ATATACATcT	3240
	GcttCGCTAA TCGTACAATA CTTTTTCCTT CTGGCGTATC ATCGTAGATT GATGACATAT	3300
25	AAGCAGCGAC TATCAATTTT TCAAGCATTT GTTGATTCAC TGGTAAAAAT TCACTAGCGA	3360
	TTCGATTGCC ATAAGTGATT GTGCCTGTCT TGTCTAAAAT CATTACATCG ACATCTCCAC	3420
30	ATACTTCTAC AGCAGGCCCA CTTTTCGCTA ATACATTGAA TTGAGTAACA CGATCCATGC	3480
	CTGCAATACC AATCGCCGAT AACAAACCAC CGATTGTCTG TGGTATTAAA CATACTGTTA	3540
	ACGCAATGAG CATCGCAATA GGTAAAATTA AATGCAGGTA AGATGCTATT GGATATAACG	3600
35	TTACAATAAC GACTAAAAAT ATAATTGTTA ACGTTGTTAA TAATGTAAAA AGTGCAATTT	3660
	CATTTGGTGT TTTATTTCTT TCCGCCCTT CAACTAAGGC AATCATTTTA TCTAAAAAAG	3720
	ATGTACnCGC TTCACTCTCA ACACGTATTT CTAACCAATC AGATGTTACA AGTGTACCGC	3780
40	CAATGACTCC ATCAAAATCG CCACCTGATT CTTTTATCAC AGGTGCAGAC TCACCAGTAA	3840
	TTGCAGATTC ATCAACGGTT GCTAATCCAT TTATTACAAC GCCATCAGCA GGGATTGTTT	3900
	CTCCATTTTC TACCCGAATA TTTTGTCCGG CTTTTAACTC TGTGGCGTTC ACTATCCGAT	3960
45	ACGCACCATT TTCTTCTATC AATCGAGCAG TTAAATTTGA TTGTGCTTGT CTTAAACTAT	4020
	CAGCTTGCGC TTTTCCACGA CCTTCAGCAA AGGCTTCTGA AAAATTAGCA AACAATATAG	4080
50	TTATTAATAA TATGATAAAA ATTGTAATCA AATAACCTCG CGATAGATAG CTAGTTCCAA	4140
	ATATGTCAGG AAAACATATT AATATCAACG TTAATCAT TCCAACCTCA ACGACAAACA	4200
	TTATCGGATT TTTTATTAAT TGTTTAAGAT TCAGCTTATA AAAACTCATT TTCAAAGCTT	4260

55

	TTTATTTTAA AGTTAAAAAT TCACCAATAG GACCAAGTAA TAGTACTGGA ATAAATGTCA	4380
	AACCACTTAG TAAAACGATA AATACGATTA GTGATACGCC AAAATAAGGT TTATCAATCG	4440
5	CTATTGTATA TTTATCTTGA TGGTATGATT TTTTATTAC TAAACTTGAT GCAATCATT	4500
	ATTGCAAAAT AATTGGTATA TAACGAGAAA GCAACATAAT GATTCCTGTA GAGATATTCC	4560
10	AGAATGTTGT ATCATCTTTC AGTCCTTCAA ACCCTGATCC ATTGTTGCA GCAGCTGATG	4620
	TCATTTCATA CATAACTTGT GAAATACCAT GAAAAGACGG ATTCGTtATa CTTtCACTTG	4680
	CTCCAGGAAT CATAAAAGCA AGTGCTGAAA ATACTAAAAT TAAAATTGGG TGTATGAGAA	4740
15	AGACTAAGAC AATACATTTT ATTTACGGG CGCCAATTGG CATATTTAAA TATTCTGGTG	4800
	TTTTACCAAC CATCAAACG CATATAAACA CCGTCAGTAA GACAAATATC AATAAATTC	4860
	TGAGTCCTAC GCCTTCGCCA CCAAATACAA CATTTAGCAT CATTAATACC ATTGGTCCTA	4920
20	ATCCACCTAT AGGCGTTAAG CTATCATGCA TGTATTAAAC AGAACCCGTT GTAAATGCCG	4980
	TCGTAATAAC TGTAATAGT GCTGACAAAC CTGCTCCAA CCGTACCTCT TTACCTTCCA	5040
	TATTGGGTCC ATAAATGCCT AAATTCGCTA GTATTGGATT ACCACGATAC TCACTCCACA	5100
25	TAGTTAATGT AAGAATTGCT ATAAAAATGA AAAACATTGC GACAAATAAT ATCAACGCAT	5160
	GACGATGTAC TCGTTTACCA TGTCTACTTA ACATGCGACC AAATAAGAAC AACATTGACA	5220
30	TAGGAAGTAA CATCATACTG CCCATTTCTA TAAAATTGCT CCAAATATTT GGATTTTCAA	5280
	AAGGTGTTGC AGAATTTCTT GCTAAAAATC CTCCACCATT CGTACCAAGA TGTTTTATTG	5340
	ATTCAAGTGA TGCAATAGGT CCAAATGCAA TATGTTGAAT ATGTCCGCTT AAAGTCCGAA	5400
35	TCATTAAATT AGCATGCAAC GTTGTGGTA CaCCTTGAGT CATCAATAAA ATACTAATTA	5460
	AACATGATAA TGTTAAAAGT ACTCGGACAA TAAACCGAAC AATATCTTGA TAAAAATTAC	5520
	CAATGATATT AGTTAATCCA GTTAAACGTC TCAACATCGC TATACAAACG GCGTAACCTG	5580
40	ATGCACTAGA TGTAACATT AAATATGTCA TTACAATCAT TTGCGTTAAA TATGTCACAT	5640
	CTGaTTCACC GTTATAGTGT TGtAAATTAC TATTTGTTAA AAAAGATATT GCTGTATTAA	5700
	ACGCTAAATC TATCGATTGG TTTAAATTAT GATTTGGATT TAAAAAAGC CATTGCTGAA	5760
45	CTATTAGCAA TACAAATGTT ATAAACCCCA TAAATCCATT AAATGCCAGA AAATGTTTGA	5820
	CATATGTTTT AGCTGACATG TGTCTAAAT CTGTGCCGAT AATTTTAAAA CACATATTTT	5880
50	CAAATCTAGT AAATATTAAA TCTACTCTTG ACGATTGCAC CAATGCTACG CGATATAGAT	5940
	ATCCACTAAA AACATACGTA ATCATAACCA TCATTGTTAG AAACAAAATT ATTTCCATGA	6000
	TAACCCTCAC TTAATATATT TCTAAAATTT TCACTACGA ATTAAGGCAT AAAATAAATA	6060

55

ACACAACAAC ATCGTAACAA CTTGTTTATG AGAGAAATnT TAATTTTCAA ACTTAGTTAT 6180
 TAAGAAAaCA TTAAGATGTG TATGCAGAAA TAAATTTTAT AGCATTTAAT TGTGAAGAAT 6240
 5 ATTATGATAT TGCTATCGAG GTGAAGGTTA TG 6272

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1978 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

AAATGATGTT TTACAATAAA TATAnAAAACG TATCAACATA TATCATCATA TTTTITAGTTT 60
 20 CAAGTGCAGC CTTTGCAATA TTCTTGTTAA GTGCGnACAT TAGTGCTCAC TCGGAACAAG 120
 TGTACGAAAT GACTGACCAT CAAATTAAGA ACAATACGAT AAATAAAGCA TACGAACATA 180
 AAGACCCAC AAACAATAGC GAACAAAGAG ATGGGAAAGT GTTCGCTTTA ATAAATTGAT 240
 25 ACATTGTCAC AACGTTATTT TGCCTATTTT TGCgmAATAG CGTTTTTTAT TACwTTTTTG 300
 CTGATsTTAA ATTTGTTATA TTTTGTTAAA GTATTATAAT GATTGAATAA ACAAATTGAA 360
 GGTAGGTTTT TTAATTGAGT AATTCTGATT TGAATATCGA AAGAATTAAC GAGTTAGCTA 420
 30 AAAAGAAAAA AGAAGTAGGA TTAACCAAG AAGAAGCAAA GGAGCAAAAC GCCTTAAGaA 480
 AAGCTTATCT TGAGAGTTTT AGAAAAGGGT TTAACAACA AATTGaAAAT ACTAAAGTAA 540
 TTGATCCAGr AGGTAATGAT GTAACACCTG AAAAAATTAA AGAGATACAA CAAAAAGAG 600
 35 ATAATAAAAA TTAAATCACA AATCTGTAAA GAATTTTCTG ACATTATAAC TTGAAATAAG 660
 TATTTTACTT ATCTTTTTAT TTTAAAATAA GTTATAATGT ATTTGATAAA ATTGAAGAAG 720
 40 GGAAGATACA CAAGATGTTT AATGAAAAAG ATCAATTAGC TGTGATACG CTACGTGCAC 780
 TAAGTATCGA CACAATCGAA AAAGCGAATT CTGGTCATCC AGGATTACCT ATGGGAGCTG 840
 CCCCAATGGC TTACACTTTG TGGACACGTC ATCTGAATTT TAATCCACAA TCTAAAGATT 900
 45 ACTTCAATAG AGACCGTTTC GTATTATCTG CAGGGCATGG TTCAGCATTa TTGTATAGCT 960
 TGTTACATGT TTCTGGTAGT TTAGAATTAG AAGAATTAAA GCAATTTAGA CAATGGGGTT 1020
 CTAaAACACC AGGTCATCCT GAATACAGAC ATACAGATGG TGTAAGAGTT ACTACCGGAC 1080
 50 CACTTGACA AGGTTTTGCT ATGTCAGTAG GATTAGCTTT ACAGAAGATC ACCTAGCAGG 1140
 gAAATTTAAT AAAGAAGGAT ATAATGTTGT AGATCATTAC ACATATGTAT TAGCTtCTGA 1200

AAGTAAATTA GTTGTTTTAT ACGATTCAAA TGATATTTCA TTAGATGGCG AATTAAACAA 1320
 AGCTTTTTTCT GAAAACACAA AAGCTCGTTT TGAAGCATAT GGTGGAATT ACTTACTAGT 1380
 5 TAAAGATGGT AATGATTTAG AAGAAATTGA TAAAGCGATT ACTACAGCTA AATCTCAAGA 1440
 AGGACCAACG ATTATTGAAG TTAAAACAAC AATCGGATTT GGTTCACCGA ATAAAGCAGG 1500
 AACTAATGGT GTTCATGGGG CACCTTTAGG TGAAGTTGAA AGAAAATTAA CATTGAAAAA 1560
 10 TTACGGTTTA GATCCTGAAA AACGTTTTAA TGTTCAGAA GAGGTATACG AAATTTTCCA 1620
 AAATACTATG TTAAAACGTG CTAATGAAGA TGAATCTCAA TGAATTTCAT TATTAGAAAA 1680
 ATATGCAGAA ACATATCCTG AATTAGCAGA AGAATTTAAA TTAGCGATTA GTGGTAAATT 1740
 15 GCCTAAAAAT TATAAGGATG AATTACCACG TTTTGAACTG GGTGATAATG GTGCATCTCG 1800
 TGCTGATTCT GGTACTGTTA TTCAAGCAAT CAGTAAACT GTCCCTTCAT TCTTTGGTGG 1860
 20 ATCAGCAGAC CTTGCTGGTT CAAACAAATC CAATGTAAAT GATGCAACTG ATTATAGTTC 1920
 TGAAACACCT GAAGGtAAAA ATGTGTGGTT TGGTGTACGT GAATTTGCTA TGGGTGCT 1978

(2) INFORMATION FOR SEQ ID NO: 142:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TAGTAGTATT TATTAAATTA TACGAAGGGA CCcAACACAG AAAATTCATT TTATTGAATT 60
 35 TTACATTTAT GTGCCAAGTT GGGAAAAATG TCTTATTTTT TCaAGTATT TAAAAGTAAA 120
 ATTACATGTT AATACGTAGT ATTAATGGCG AGACTCCTGA GGGAGCAGTG CCAGTCGAAG 180
 40 ACCGAGGCTG AGACGGCACC CTAGGAAAGC GAAGCCATTC AATACGAAGT ATTGTATAAA 240
 TAGAGAACAG CAGTAAGATA TTTTCTAATT GAAAATTATC TTAGTGTCTGT TTTTtaggga 300
 TTTATGTCCC AACCTTTTTA GAATATTAAA TTTCTACAAT TTCGTCTCT TCAACAATAA 360
 45 AGCCCATGTG ATTGACGCTG TTATTTAAGA AAGTCAGAAT ATAACGCATT ACTTCATCAC 420
 GTTCTGGCTC ATTGTGAACC TCGTGGTAAA AACCTTGCCA AGCTTTAAAA TATAATTcag 480
 GTGTTTGATA TTTTCTTTA AACTCATCAA TTGCCCTAGT ATCAACAATT AAATCCTTCG 540
 50 TTCCATACAT TAATAGCGTT GGCATTGGTT GAATGTCATG AATATGAGCC ATCGTATCTT 600
 TCATCGTCTC ATTAATTGTA TTATACCAAT GATACGTTGC TTTTtTTAAC ATTAAccat 660

55

	CATTA AAAACG TGTGTCTTTT GAAATTTTAC CTATATTTGA AACAAAGTTTA TCTTTACGAT	780
	TTTTTCCATT CTTTTGAAGT TCTAGCATAG GAGAAATTAA CATCATCCCC TCGATTGGCA	840
5	ATTCTACTTT TTCAAGTAAA TTTAATAAAA TCAAACCGCC AAGTCCTACC CCTAATACAT	900
	AAGTAGGAAT TTTATATTCA TTAGCTATCT TTAACCAGTC TAGCAAACCTT TCGTGATACG	960
	TTTGAAAGTT TTCAATTTGT CCTTTATTAG CTCTTGAAGT TTGACCTTGA CCAGGCAAAT	1020
10	CTCCCAT AAT CACATGATAG CCATTTCTTC TTAACATCGT AATAACATAT GCATATCTTC	1080
	CCGTATGTTT TAATATATTA TGAGCAATAA CAACGACGCC TTTCGCATCA TTTTCAGCTT	1140
	CCCACCTCCA CATTATTATA CTGCCCCCTT TTCATTAATC TTCAATAACA TAATTATAGC	1200
15	AAATTCAC TA TGTAGATTTC TATTTATAGT ATTATTGTTG TCCATATTAT TATATATAAA	1260
	TGAAATCAAC ATCAATAATA GTGTAATTAT ACATAATTAT TTTTGATTGT TTTTGATGAA	1320
20	AACGCTTTCT CGAATATTTT TTTCATGCTA AACTTATTGT AAACACAAGG GTTTGGAGGA	1380
	GTAGCAATGG CACTATTAAA GAATTTTTTT ATCGGATTAT CTAATAATAG TTTTTTAAAC	1440
	AACGCAGCAA AAAAAGTGGG CCCACGTTTG GCGGCCAATA AAGTCGTTGC CGGAAATACA	1500
25	ATTCAGAGT TAATTAATAC AATCGAATAC TTAAATGACA AGAATATCGC TGTTACGGTA	1560
	GACAATTTAG GGAATTTGT CGGTACAGTT GAAGAAAGTA ATCATGCTAA AGAACAAAT	1620
	TTAACAATTA TGGACGCGCT TCATCAACAT GCGGTAAAGG CACATATGTC TGTTAAATTG	1680
30	AGTCAGTTAG GTGCAGAATT CGACTTAGAA TTAGCTTACC AAAATTTAAG AGAGATTTTA	1740
	CTTAAAGCAA ATACTTACAA CAATATGCAT ATAAATATTG ATACTGAAAA ATATGCTAGC	1800
	CTGCAACAAA TTGTTCAAGT TTTAGATCGC TTAAAGGCG AATTTAGAAA TGTTGGTACT	1860
35	GTAATTC AAG CATATTTATA CGATAGCCAC GAATTAGTTG ATAAGTACCA AGATTTACGA	1920
	TTACGTTTGG TTAAAGGTGC ATATAAAGAA AACGAATCAA TTGCATTTC AATCTAAGGAA	1980
40	GACGTAGATG CAAATTACAT CAAAATAATT GAACAACGTT TGTAAACGC ACGCAATTTT	2040
	ACTTCAATTG CAACACATGA CCATCGCATC ATTAATCATG TAAAACAATT TATGAAAGAA	2100
	AATCACATTG AAAAAGATCG TATGGAATTC CAAATGCTCT ATGGTTTTAG ATCAGAGTTA	2160
45	GCAGAAGAAA TCGCAAATGA AGGCTATAAT TTCACTATTT ATGTACCTTA TGGCGATGAT	2220
	TGGTTTGCGT ATTTTATGAG AAGATTAGCA GAACGCCAC AAAACCTATC TCTTGCTGTA	2280
	AAAGAATTTG TGAAACCTGC TGGCTTAAAA CGTGTTGGCA TAATTGCAGC TTTAGGAGCT	2340
50	ACAGTTATGT TAGGTTTAAG TACAATTAAA AAATTATGCC GTAAATAGAG CAAGACATAA	2400
	ACAATAATTT AGGAGTCTGG AACAATAATC AATGTTCTAG GCTCCTAAAT GTTATATTGG	2460

55

	TAGATTTTAA TAAATTAGCC ATTTCAATTG CACTTACTGC TGCTTCAGCA CCTTTATTGC	2580
	CAGCTTTCGT ACCTGCTCTT TCCACAGCTT GTTCAATAcT TTCAGTCGTT AAAATACCAA	2640
5	ATATGACTGG TACATTAGTT TGATCATTCA CTTTAGAAAC ACCTTTCGCG ACTTCATTAC	2700
	AAACATAATC ATAATGAGAC GTAGACCCGC GAATTACGCA TCCTAATGTA ATTACTGCAT	2760
	CATAATTTCC TGATGAGGCT AATTTTTTAG CTACTAAAGG AATTTCAAAC GCACCTGGCA	2820
10	CAAATGCTAC ATCAATATTG TCTTCATTAA CATCATGTCG AATCAAAGTA TCTTTTGCAC	2880
	CTTCAAGTAA TCTTCCAGTG ATAAAATCAT TAAATCGACT AACTACGATT GCAACTTTCA	2940
	AATCTTTTCC AATTAATTTA CCTTCAAAAT TCATGTTAAA ATCCTCCTAT ATTAAATGAC	3000
15	CCATTTTTAT TTTTTCGTT TCCATATAAT CATGATTATG TACCGTTTCT GGTACGATAA	3060
	CTTCAATTCT TTCTGCAATA TCAATGCCAT ATTGTTTTAA TCCCTCAAAT TTA CTGAT	3120
20	TATTACTTAA TAAATTGATA TGTTGATGT TAAAATATTT TAAAATCTGT GCAGCAATAT	3180
	GATAATCTCG CAAATCTTCA TCAAAACCTA ATGCTAAATT TGCAGTTACT GTATCATATC	3240
	CTTGCTCAAT TAATTCATAT GCGCGTAATT TGTTTAACAA TCCTATGCCA CGACCTTCTT	3300
25	GAGGTAGATA AATAATCATG CCACCATGTT CATTGATATA CTTCATAGAC GATTCAAGTT	3360
	GAGCACCACA ATCACAACGT TGA CTATGGA AAATATCGCC TGTAAGgCAC GCAGAATGTA	3420
	AGCGTACATT TTCATGTTGT CGAATTGCAC CTTTTGTCAG TACAACTATC TCTTCATCTG	3480
30	TGTATGTCGC TTTAAACCA TACATATCAA ATGTTCCGAA ATCTGTAGGC ATTTTCACTT	3540
	TTGCCTTAAA TTCAATTTCT GGTTCCTAATT TTTTACGATA TTCAATTAAA TCATCAATCG	3600
	TAATCATCTT TAATTGATGT TTTTCTTTAA ACTTTTGTA ATCTTGTCCT TTCGCCATCG	3660
35	TGCCGTCATC ATTCATAATC TCACAAATGA CACCAGCGGG CTTGGCACCA GTAAGTTTAG	3720
	CTAAATCAAC AGCCGCTTCT GTGTGTCCAT TTCTAGCTAA TACGCCTTTA TCTGTGCTA	3780
40	CTAATGGAAA TAAATGACCA GGACGATTAA AATCTTTAGC TTCACTACTA GGATCAATGA	3840
	GCTTTTTGGC AGTCAATGTA CGTTCATAAG CACTAATTCC TGTGTTGTA TCTACATGAT	3900
	CAATACTCAC TGTA AATTGC GTACCAAAGA TGTCGGAGTT ATCATCAACC ATTTGTACCA	3960
45	AATCCAAACG TTGTGCAATA TCTTTAGACA CTGGTGCGCA TATTAATCCC CcTGCTTCTT	4020
	TCGCCATAAA ATTAATGGTA TTATCGTTCA TCCATTCACT AACCCTACT AAATCACCTT	4080
	CATTTTCACG ATTCTCATCA TCTACTACAA TAATTGGTTC TCCATTTTTT AAAGCCATTA	4140
50	AAGCACTGTC AATATTATCG AATTGCATGC TACCCTCCT AAAAACCAA TGCTCTTAAT	4200
	TTATCTACAG ATAATTGGTC TTTATCTTTA TTTAAAATAT TTTCAACATA TTTAAACAA	4260

55

	CTCGTTTCTG GAATAAGATG AATGTCAAAA CTGTTATCAT GCTTATCAA TACCGTTAGA	4380
	CTAACACCAT CCACAGTAAT AGACCCTTGC TTAACCTAAT GATTATTAAT ATGTTGGCTA	4440
5	CATTGAATCG TAATAATTTT TGCATTGGCT GTTTCATTTA TTTTGAAC TGTTCCCTAGT	4500
	TCATCTACAT GACCGAGGAC AAAATGTCCA CCAAACCTAC CGTTACCACT CATGGCACGC	4560
	TCTAAATTTA CTTCTGATTG TCGCTTAACA TCTGCTAAAT AGGTTTTATT TTCAGTGCCT	4620
10	TTAATTACTT GAACAGTAAA AGATGTCTGA TTAAAATCAA TCACTGTTAA ACATGCACCA	4680
	TTAACTACTG TGGAATCACC AATATGCATA TCTGCCGTAA TCTTATGTGC TTCAATTTCA	4740
	ATCGTCCTGA CTGATTGACG AATTTGAACA CTTTTAACGA CACCTATTTT TTCAACGATG	4800
15	CCAGTAAACA TGCATCATCA CTTCTTTTCGT AAAGTTAATT TAACATTTTG ATTTAATAAC	4860
	TCGGAATGAA CAATTTCAA TTGGTTTCGCA TCTGGTATCT CAATCACATC ATTTGTTTGA	4920
20	TAAAATTGAT AATTTCCAGA TCCGCCAATT AATTTCGGGG CATAATAGAG AATAAATTCA	4980
	TCTATATAAT TAGATTGGAG AAATTCTGAA GTAGTGGTTG GACCTGCCTC GACTAGCAAA	5040
	GTTCCAACCTC CTCTTTTATA TAAATTGTGA AGAATTGTTG TTAAATCGCA AGACTTCAAG	5100
25	TAAATAATTT CAATATGTGT TTGATTGGTT GTTAAATTG GATTTTCAGT ATATATCCAA	5160
	ATTGGTGTTG ATTCATCTTG ATAAATTTGC TGATTAAAT GAATATTCCC AGACTTAGAC	5220
	AATATTACTT TTATAGGGTT TTTTCCATCT TGAATACGTG TAGTATATTG TGGATCATCT	5280
30	AATTCAACTG TACGTCTTCC AGTTAACT GCGTCGTGTC GATGTCTTAA CTTATAGACA	5340
	TCTTGTTTAA CCTCTTTGTT AGTAATCCAT TGACTTTGTC CATTATCATT CGCTTGTTTA	5400
	CCATCTAAAC TTGCAGATAC TTCACTGTA ATTTGTGGCA GTTGCTTTGC TTTTGCTTTA	5460
35	AAAAAGTCTT GGTATAATTG TGATGCCCCG TCATCATCAA CGCATTCAAC CTCAATACCG	5520
	TGAGCCCGTA ACGTCTCATC ACCATGTGTC TCTAACGAAT TGTCTTTTGT TGCGTATACT	5580
40	ACTTTTGCTA TCTTACAATC AATTATTTTG TTAACACAGG GTGGTGTTGA ACCAAAATGA	5640
	CTACATGGCT CTAACGTAAT ATAAATCGTC GCACCTTCAG CATTTTGTTG TGCCATATCA	5700
	AGTGCTTGAA CCTCCGCATG CTTGTCACCT TTTCTCAAGT GTGCACCAAT ACCAACAATC	5760
45	CTACCTTCTT TAACTACAAC AGCGCCAACG GGTGGATTAA CACCTGTTTG ACCTTGTTACC	5820
	ATATTTGCAA GTTGAATCGC ATAATCCATA AATTGACTCA AATGATCACC TCTATAAACA	5880
	AAAATCCTCA CATCATGAAT TAAGATGCAA GGAGAAAAAT TTATCGTTAA ATAAGCCTAT	5940
50	TTGTACACAT TTTTACAAAT ACGCTACATT ATCTTTGTG ATAATTAACA TTCTTTCTCC	6000
	CATCCAGACT TTAACGTGTC GCTCTAGAAT CTCACTAGAT CAGCCACTAA TATGAAACAT	6060

55

TTaTATATGA AATTGTTATA GATTATTTGA GTACGTAGTA TGTCAACTAC ATTTAAAATG 6180
 ATACTATATG TTTTCTGAAA AAACAATTAA TGACGGTTTT AATTTAATAT AATCTGAGTA 6240
 5 CTATAGGCAT CTCATTGATA TGATTCTTAC TAACAGACAT TAAAATCAAA CCTTCAATTC 6300
 GTCTCTATAG AGCGTTCTCT TTATTATCTT CTAGTTACAA ATTATTGATT GtCACTGCGC 6360
 TGTGTGTGCT CATTCGATTC TAAAGCATCA TATAATTGAG ATACTGTATG CGCAACTTGT 6420
 10 TCTACAATCA TTTTCACACC GTTTCGTAGT TTATTAACAC CGTTTGTCAT TTGACCTATC 6480
 GCAATCATAT TTGTTAATGT TCCAAACCTT GGACTAATAA CTTGATTGGT TTCCGGAATG 6540
 ATTTGTATGC CTCCCATTGG GTGTGCTTGT ACAATTGTC TATTTTCAAG ATTTCTAATT 6600
 15 AATTGATCAT CTTGATCCAA TTCATTTAAA TGACTTTTTG CACCTGTCGC GTTAATGACA 6660
 ACATTATATA TGTCTACTGA TTCTTGTTTT TTGTATGAAA AATAATACAA CTTGCCATaC 6720
 ATGTTACAT CTTCTAAATC TTTTTCAAA ATTAAAGACT TATTTTCTAT TAATTCAATA 6780
 20 ATTAGTTCAG CAGTTCTTGG AGGCATTGGA TTTGAATTTA ATTGAATCAT CTTTGAGTAT 6840
 TTTTGATTAA ATTGATGTTG GTCTTCAATA CTTAAGCTAT TCCATATCCA ATTTAAATTC 6900
 25 TCTTTCAAAT GTTCAATCAT ACTTTGGAAA ATGCCCaTTT CTGTTGGACG CGCTAAATCA 6960
 TACTTCAAAT CTGCAATATG ATTTCTGTGA CGTCTATGTA CTAATTTTTT AAAATCAATG 7020
 TCATATTCAG CACATTCTTT TAAAAATAAA GAACTAAAG TATCAAGCGG TGCATTGCCG 7080
 30 AAATGATGTT TTTTAATGTC ATTTAATTTG TCTTTAGTTA AGTACTTGAA TGTCACGTCT 7140
 ATCATGTAC CTCTTACACT TGGTAAATGA GCAGAACGAC TCGTCATAGT AATTGGTAAT 7200
 TTTGGATGAT GAGCAGCAAC ATAACGGACA ACATCTAAAC TGGCAAGGCC TGTACCAATA 7260
 35 ATCGCAATAT CGTCCAGTTC ATTTACTTCG TCTAACGTAT TATATGTTGG ATAAGGCGTA 7320
 gcGATATATC CTTTTTTACC CTTTAAGTTA TATGGATCAT GGTAGGCAAA TGTACCACAT 7380
 GTTAAAAATA CATAATCGTA CGCTTGCCAT GATTGTCCTG AATTTGTAGT ACATATGTAA 7440
 40 TAAGTTAAAT TCGTTTCATC GATATTAGAA TTTGTATAAA TCTCTGAAC TTTATTATAA 7500
 TTAGTTGATA TATTTGGATA TTTTTCGTG AACATAGATA AATAAGATTT CATATAATGT 7560
 45 CCGAATACAA ATCTCGGTAA ATATGCAG 7588

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10320 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

	nCTAGGTATT TTAAACCTAA TCTAGATAAA CTAGCTTCGT AAGCAGCTGC TACATTTTCA	60
5	CGACCGAAAT CCTCAAAATA TAATTTTGAA GTAATAAATA AGTCTTCTCT AGCAATACCA	120
	GTTGACTCCA ATCCGGCACG AATGCCAGCA CCTACTTGTT CTTCAATCCC ATAAACTTTT	180
	GCGGTATCAA TACTACGATA TCCTTGTTCA ATGGCATACT TAACACTTTC CATGCAATTT	240
10	TCATCATTTT CCACACGAAA TGTCCCTAAA CCAATTTGTG GCATCGTGTT TCCATTATAA	300
	AATGTTTTAA CCTCCATAAA TATCGCCTCA CCTTTTGAT GTATTATACC CTGTTATCAT	360
15	AACAAATCTG AGTTGAATAC ATGAGAAAAA ACACTTAGAG CAATCAACCA CTAAAATTCT	420
	AGTAATATCT CTCAAATATT AATCAAATTG TAAAAGTAAT TCTGTTTAAT TTATGACAAA	480
	CTAAAAAAGC CGAAGTAACA ACATATAGTC ATCACTTCAG CCTAACATTT AATTGAATGA	540
20	TTCAATTTTA TCCATCATTT GTTGTAAGTC TTCCACGTTG TATTGAATAC GACCATGGAA	600
	TACAAATTTG TTAAAGAACT CGTCTAATTG TTCAGCACCG ACAAGCACTT TGACAGCACT	660
	ATTTTGATTA TAATTTGAAA TCGTTACATC GCCTTCATTT TTAAGATTAA AGTATAAAAT	720
25	TGAAGTTGGT GTATATTTGG CACCTAATTC TTTTGTAAAG TCTTCAGCCA ATTGTTTAAT	780
	CGCCTCAATT TGATCTGAAT AATTTACAAA TGATAATGAA CGTTTGTCAT CATTTTGATC	840
	CATCACAATA GTTTGCGGTC TAGATTTATC TAAATCCAAT GTATCAAATA CTTGTTCCAT	900
30	TGGTGGTAAA TCTTTAAATT GACCGCCACT AATACCATT AAAACATGAC CTTTTAACAA	960
	TTGAGAATCA ATAATATAAA GACCAGTTCT TGTTAATACT AAATGACTAA TTCGTTCAAT	1020
	ATTATTAAAG CCATCCTTTG GTAAAAAGAT ATTTGCCATA ATGTGCATAT CTTCTGGTCG	1080
35	AATTCGTTTT TCTTTAACTA ATCTTTCACG AATACCAATT AATCTCATGT CCGTTACATA	1140
	TTCACTATGA TTTTTCGAGA ACAATTTTAA TGCCTCAATC TCACGATCTT TTGTTACTAAC	1200
40	CATGTGATTA TAATCTTCTT GTTGTTTTGT AATTGTCTTT TTATTTTGAA TACGCTCTTT	1260
	CTCTAAAGCT TCTTCATGAG ACTTTTAAAT GTTTTGTTCT TGTGTTTCAT ACTTTTCTTC	1320
	TGTTTGTCGC TTAACTTTTT TCTTACTACC TAAGGCAACT AAAAAAGGA CAAAAAGAT	1380
45	TAATGCAATG AgCTACTGCA ATAATGAGTC CAATGACTAT CGGTGAAGAT AAATCCATCA	1440
	CAACAACGCT CCTTTTAAAT ATATGAATAA CTTTAATTAT AATAGAAAG CTAAAGATTT	1500
	TCGATACATA TTATCATTTA TATACCGAAA ATCTTTTATT TAGCTATATT CAATTCATCT	1560
50	TATTATTTTA CTGCGTCTTT TAATCTTCC ACTTTGTCTA ATTTTCCCA TGGGAATAAG	1620
	ACATCTGTAC GTCCAAAATG ACCATAAGCA GCAGTTTGTT TGTAATCGG TTGTTTCAAA	1680

55

	AGTTGCCCTT CAGAACTTT ACCTGTTCCA AATGTATCAA TTGCAATTGA CACTGGTTCT	1800
	GCAACACCAA TCGCATATGC CAATTGTACT TCACATTGAT CTGCTAAACC TGCTGCAACA	1860
5	ATATTTTATAG CCACATAACG TGCAGCGTAT GCAGCTGAAC GGTCTACTTT TGTAGGATCC	1920
	TTACCACTGA AGCATCCGCC ACCATGACGT GCATAGCCAC CGTACGTATC AACAATGATT	1980
	TTACGTCCTG TTAATCCTGC ATCACCTTGA GGTCCACCGA TTACAAAGCG TCCTGTAGGA	2040
10	TTGATGTAGA ATTTAGTTTG TTCATTAAATC AAGTTTCTG GAACAGTTGG ATAAATGACA	2100
	TGTGCTTTAA TGTCTTCTG AATTTGTTCA AGTGTACAT CCTCAGCATG TTGTGTTGAT	2160
	ACGACAATCG TATCAATACG TACTGGGTTA TCATTTTCAT CATATTCAAC AGTGACCTGA	2220
15	ACTTTACCGT CTGGTCGTAA ATAATTTAAC GTACCATCTT TACGCACATC TGATAAACGT	2280
	TTTGCCAATT GATGTGATAA ATAAATTGCT AGAGGCATAT ACGTCTCTGT TTCATTCTGT	2340
20	GCGTAACCAA ACATTAAACC TTGGTCACCT GCACCTGTTG CTTCAATTTT TTCTTCGCTA	2400
	TCTTTATCAC GATACTCTAA TGCTTTATCC ACGCCTGTG CAATGTCAGG TGATTGTTCA	2460
	TCAATCGCAG TTAAATTCG CATTTGTTCA TAATCATAAC CATATTTTGC TCTTGTTAT	2520
25	CCAATTTCTT TAATTGTTTC TCTAACAACT TTCGGAATAT CAACATATGT TGTGTAGAA	2580
	ATTCGCCGG CGATCAATGC CATACCTGTT GTAACAGTTG TTtCACAAGC TACACGTGCA	2640
	TTTGATCGT CTTTTAAAT AGCATCTAAT ATTGCATCTG ACACTTGGTC AGCGATTTTA	2700
30	TCTGGGTGTC CTTCTGTAAC AGACTCTGAA GTAAATAATC GTTTGTTATT TAACATAGTT	2760
	TGCTCCTTTA AATTTATATT ACGAAAATTC TCTCTCTGTG AGCTAAATAA AAAAGACCTT	2820
	CTAACTATTA ATATAGAGAG AAGGCCTAAT ACGTCCATTC GCTCTTATCG TTCAGACCTA	2880
35	TTTGCTGCA AAcGGTTTG CACCTTTCTT TTATAAAAAA GAGGTTGCTG GGTTCATTG	2940
	GGTCCATGTC CCTCCACCAC TCAGGATAAG AGAATCCGTT AAAAATAATA GTACCTAATT	3000
40	AATGAATTAA TGTCAATTTT TCACAAATAA ATTTACAGTA AAATATTGTA GATTAATTAT	3060
	GTTAATGTGT TATACTAATT AAATGTAAAG GCTTACATTT AAATTATCGC TTTGGAGGGA	3120
	TTTAGGATGT CAGTAGACAC ATACACTGAA ACAACTAAAA TTGACAAATT ACTGAAAAAA	3180
45	CCAACGTCAC ATTTTCAACT TTCGACGACA CAACTTTATA ATAAAATCTT AGACAATAAC	3240
	GAAGGGGTAT TAACAGAACT TGGTGCTGTT AATGCAAGTA CTGGAAAATA TACTGGTCTG	3300
	TCGCCTAAAG ACAAATTTTT TGTCTCTGAA CCTTCATATA GAGATAACAT TGATTGGGGA	3360
50	GAAATTAATC AACCTATCGA TGAAGAACT TTCTTGAAGT TATACCATAA AGTACTAGAC	3420
	TATTTAGATA AAAAAGATGA ACTATACGTA TTTAAAGcT ACGCTGGTAG CGATAAGAT	3480

55

	ATGTTTATTA GACCTGAATC AAAAGAAGAA GCTACAAAGA TTAAACCTAA CTTCACATATC	3600
	GTTTCTGCAC CACATTTTAA AGCAGATCCA GAAGTTGATG GTACTAAATC TGAAACCTTT	3660
5	GTCATTATTT CATTTAAACA CAAAGTCATT TTAATCGGCG GTACTGAATA CGCTGGTGAA	3720
	ATGAAAAAAG GTATCTTCTC TGTAATGAAT TATCTCTTAC CGATGCAAGA TATTATGAGC	3780
	ATGCATTGCT CAGCAAACGT TGGTGAAAAA GCGGATGTTG CATTATTCTT TGGTCTATCT	3840
10	GGCACTGGTA AAACAACCTT ATCGGCTGAC CCACACCGTA AACTAATCGG TGATGATGAA	3900
	CACGGCTGGA ATAAAAACGG GGTCTTTAAT ATCGAAGGTG GCTGCTATGC AAAAGCAATT	3960
15	AATCTTTCCA AAGAAAAAGA ACCACAGATT TTTGACGCAA TCAAATATGG TGCAATTTTA	4020
	GAGAACTG TAGTTGCAGA AGATGGTTCA GTGGACTTTG AAGACAATCG TTATACAGAA	4080
	AACACGCGTG CCGCTTATCC AATTAATCAC ATTGACAATA TTGTAGTACC ATCTAAAGCA	4140
20	GCACATCCAA ATACAATTAT TTTCTTAACT GCGGATGCAT TTGGTGTTAT TCCACCGATT	4200
	TCAAAGTTAA ATAAAGACCA AGCAATGTAT CATTTCTTGA GTGGTTTCAC TTCTAAATTA	4260
	GCTGGTACAa GCGTGGTGTG ACAGAACCTG AACCATCATT CTCAACATGT TTCGGAGCAC	4320
25	CGTTCTTCCC GTTACACCCT ACTGTTTACG CTGATCTATT AGGTGAACTT ATCGATTTAC	4380
	ATGATGTTGA TGTTTATCTT GTTAATACTG GATGGACTGG CGGAAAATAT GGTGTAGGAC	4440
	GTAGAATCAG CTTACATTAC ACACGTCAAA TGGTAAACCA AGCGATTTCT GGCAAATTGA	4500
30	AAAATGCAGA ATATACAAAA GATAGTACGT TTGGTTTAAG CATTCCTGTA GAAATTGAAG	4560
	ATGTACCGAA AACAAATTTTA AATCCAATTA ATGCTTGGAG CGACAAAGAG AAATATAAAG	4620
	CACAAGCAGA AGATTTAATT CAACGTTTTG AAAAGAACTT CGAAAAATTT GGTGAAAAAG	4680
35	TTGAACATAT TGCTGAAAAA GGTAGCTTCA ACAAATAAAT TTGAATACTA AATCaAAACC	4740
	ACCGGTGTGA ACGGGTGGTT TGTCTGCGG CTATAAGCCT TCCTTACTGG CCAGCCCTAA	4800
40	AAGGGCACTG ACAAGTCAGC CAACTGCACT ACTATTCCAG CAACCCTAAA GGGTTACTCT	4860
	TTTTCTTTT TTTTTTTATT TTTCTCTCCA GTGAAAGGAT CTAAATATTC TTCCATTGAG	4920
	ATTTGGTCTG CAACGATATC CTCTTGTAAT TGATTACGAA TATAATTTTC AATCACTTTT	4980
45	TTATTTCTAC CTACTGTATC CACATAAAAT CCTTTACACC AAACTTTTCT ATTTCCATAT	5040
	CTATACTTTA AGTTAGCATG TCTATCAAAT ATCATTAAAC TACTTTTTCC TTTTAAATAG	5100
	CCAACAAATG ATGATACCCC AAGTTTGGGT GGTATACTAA CTAACATATG GATATGATCT	5160
50	TTACATGCCT CTGCTTCAAT TATCTCTACA CCTTTTCTTT CACATAATTG ACGCAATATA	5220
	ATCCCTATAT CTTTTTTTAT TTTTCCATAT ATCACTTGTC TTCTGTATTT AGGTGCAAAG	5280

55

	AAATAGCATC TCCTCGTGTT GATTATTTTG GTTGGCTGAC CAATATTTAT TCTAGCACGT	5400
	AGAGATGCAT TTTTGTGAC AATGGTAGAA CCTTTTctGa ACCATACGCA TAGCGTATGG	5460
5	TTTTCTTTTT ACAATTAAAG AGCCAACCGT TGTATAGTC TAACAATGGT TGGCTCCTCT	5520
	TATTTTATGT GCTAAAAATT TATAGGCAAT TTTATTACAA CAATGTACAT TTAAGGTGAC	5580
	CTTCATGCCA AAATCGCATC ACTCATTTAA TGGAAGCAGC ACGTCTTCAT ATAAAGTACC	5640
10	GATCCCTAAT TCAACGCATG TAGTACCACA TCTTCAAAGC TTGATAGTTC CCATGCGCAC	5700
	ACCACGTTTC ATACTAGCTA TGCGACTCAA CTTGGTTCAT AAACCTCTTTA ATATAAGTCA	5760
15	ATGTTTCAAC CATCGCTGGT GGTCTTGGCA CATGTCTTC TGCCATTTGA TAAAAATGTTT	5820
	CATGCGTGGC ACCTTTTAAC TCTAGTTGGT CCGCTAAATA ATACGCATGA TGAATACCAA	5880
	CTTGCTGGTC TTTCCCTCCA TGTACAATTA ATATTGGCGG ACTGTTTTCA TTAATGTTTG	5940
20	GAATCGCTTG GCGTGCCTCA TATGCCGCTC GATCTTTTTT CGGATGACCA ATCATTCTTC	6000
	GTAGCATGCC TCTTAAATCG ACACGTTCTT CATAATTAA ATCAATATCT GAGACACCAC	6060
	CCCAGATTGT ATAACTTGTT ACTGGTAAGT CTTGAAATGT CAACAATCCT TGTAACCAC	6120
25	CTCGCGAAAA ACCAACCATG TGGATAAATG CATGTGGATA TTTATCATGT AGCAACCTTA	6180
	ATAATTGCGT CACATCATTT AAATCGCCAC GGTAAAATTC GTCTTTGCCT TCACTCCCAT	6240
	TGTTACCTCG GTAGTATGGC CCAATCACTA AAGTTTGAAT ATCTGAAAAT TGCATTAATC	6300
30	TACCTGCGCG CACACGTCCT ACTTGACCTT TGCCACCTCG CAAATAAACT ACAATGCGAT	6360
	TTACTTCATG ATGTGGTGTC ATCATTAAAG CTTTTACTTG TAAGTCATCT GACAAATATG	6420
	TAATTTCTTC GAATTGATGC GTAAAATATT CAATTGGCAT TCGTTTACGT TTGATAAAAC	6480
35	CCAAGTGATT GCACCCTCTC TACGCATTTT AAAATGGTAC TATCTTGCAG TAAGAACTC	6540
	CGTTGTGCGA GTTCAATATC ATTGATACAG TTAACAACA CTGGCCCTGC TGTTTCTAAA	6600
40	TAATCGTTCT TGCTTACCAA TGATTCAACT TCGATAAAAT ATACATCTTT TACAAAATCA	6660
	GTTTGATCAT GTGTTTCAAT GGTATATTGT GCTATGTAAT AAATATTTTT AACTTTGGCG	6720
	CCTGTTTCTT CATATAATTC aCGTGTAAT GCTTCAGCAC TACTTTC CCC GCGTTCCTT	6780
45	TTACCACCAG GAAATTCAAT CCCCCGTAAA TTATGTTTGG TAAAAAGCAA TTGATTTTTTA	6840
	AACGTTGGAA TAGCTAGCAC ATGATTGCCA TCTGCTATCT CATTATCCTT TTTAAATGTC	6900
	AAATTAACCTT GACGATTATC TTTATCCCTA AACTTCACGC GCATCACATC CCTACATTGT	6960
50	ATGTTAATAT AATAGTTAAT TACTATCGTT GGAGGCATTA ATTATGAAAA AGATATTCTT	7020
	GGCGATGATT CATTTTTATC AACGTTTCAT TTCGCCACTC ACTCCACCAA CTGTGCGTTT	7080

55

	CCTTTATTTA	GGTATCCGTC	GTATTTTAAA	ATGTCATCCG	CTTCATAAAG	GCGGCTTTGA	7200
	CCCTGTTCCG	TTAAAAAAG	ACAAGTCAGC	AAGCAAGCAT	TCACATAAAC	ATAACCATTAA	7260
5	ATATGGTTGT	AATTGAGTTA	TATCCACTAA	AGGGGGGCGA	AATTCGAGTC	GCCCCCTCTTT	7320
	TAATATGCCT	GAATGCGCCA	CCACATCTTG	TTCAAAATAA	TAACCTGCTG	GTGTAACATC	7380
	TCCTGGATAA	TCACCTTTAC	GAGCAAGCAT	CGCTGTAAAA	TAGCGGCTTA	AACCATATTC	7440
10	GTACATGCCG	CCAATAACCA	CTTTTGCACC	ATGACTTTTC	AAAGTATCAA	TTGCCGTTTG	7500
	CACTTTATCA	ATGCCACCTA	GACGAAATGG	TTTTAATACA	ACAACTTTCA	CATTGTATAA	7560
	TTCTATCAAA	TTAATTATGT	CCaACAACGA	TGTTGCCTTT	TCATCAAGGG	CTATTGGAGG	7620
15	TATTGTTCCA	TCCGCTACTT	CATCAAGCAT	GGAGATATCT	TTAAATGGCT	CTTCGATATA	7680
	AAGAACCTGT	TCACGCGCTA	ATAACTGTAA	CTGTGTGAAA	TCTTGACGAT	CCAAGGACTC	7740
20	ATTTGCATCT	ATAACCAATT	GAAAGTGAAA	GTCTAATTCC	CGTAACACTC	TAATTTGATG	7800
	CATGATTTGA	GGCGTCCATT	TTAATTTAAT	TCTGGTCGGC	TTTGTTGCTT	TTAATGACTC	7860
	TAGTTGTTTA	TTTGATAAGC	CGCTCGcTGT	CGCTCCATAT	GCTACTGAAA	ATGAAGGCAG	7920
25	TACATGAAAC	ATTTGATACA	ATGCCATGAC	AATAGTTGCC	CTTGCAGCAG	GCGTATTTTC	7980
	CAATGAATCT	ACTAATTTTA	GTGCTGCTTC	ATACGTTTCA	AATGATTTAT	TTCTATTATC	8040
	TTCGAACCAT	TGCTCAATTA	CATGTTTCAC	TGAGGCAATT	GTTTCATGAT	CATACCAATC	8100
30	TGTTTGAAAA	GCGTTACATT	CCCCGAAATA	TGCATTTCCCT	TTGTCATCAA	TCAATTCGAT	8160
	AAACAAACAA	TCACGATGCG	TTAAAGTGAC	TTTCGGTGTT	ACAATTTGTG	ACTTAAATGG	8220
	CTCACTATAT	TTATAAAAAT	GCAAAGCTGT	CAACTTCATC	AAATCATCCT	CTATACAACT	8280
35	TATTTCTTTG	TAATTTACCT	GTTGATGTAT	AAGGTAAAGT	ATCAACCTTT	TCAAAGTGTT	8340
	TCGGTACTTT	ATATTTTCGCT	AAATGTTGTG	ATAAATATGC	AATCAATTGT	GCCTTTGAAA	8400
40	TGTCACTTTC	ACTGACAAAA	TATAATTTAG	GCACTTGCC	CCAAGTATCA	TCAGGATGCC	8460
	CTACACATAC	TGCGTCACTG	ATACCTGGAA	ATTGctTCGC	TACCGTTTCA	ATTTGATATG	8520
	GATAAATATT	TTCAACGCCA	CTAATAATTA	AATCTTTACG	TCGGTCATAA	ATCATGACAT	8580
45	AACCTTCATG	ATCTATTTCA	GCAATGTCAC	CCGTATTAAA	ATAACCATTT	TCAAACGTAC	8640
	CCGTAAATC	TGTTGGATAC	AAATATACAT	TCATCACATT	GGCGCCTTTA	ATCATTAAAT	8700
	CTCCATGACC	TTCTTTATTA	GGATTTTTAA	TTTTTACGTC	AACATTGGCA	CTTGGCATCC	8760
50	CTACAGTGTC	AGGACGTGCA	TGCAACATTT	CCGGTGTTGC	TGTTAAAAAT	TGCGAACATG	8820
	TCTCAGTCAT	ACCAAATGAA	TTATAAATTG	GCAGGTTATA	TTGTAATGCC	GTCTCTATCA	8880

55

AACCTTGTTG CATAAGCCAA TTTAAAGTTT GTGGCACAAG CGAAATGTGC GTGATTCGTT 9000
 CATTTTTAAAT CATCGTTAAA ATTTGTTCCG CATTGAATTT ATCAACAATG CGCACAGTAA 9060
 5 AACCTTCAAT AACAGCTCTT AAAAGTACAC TGAGACCCGA AATATGATAA ATCGGCAAGA 9120
 CAGATAGCCA ATTAGTGTC AATCAAATC CCAAGCTCTC TTTACATCCG ATTGCACTGG 9180
 CATAATGATT ACGAAACGTT TGTGGCACCG CTTTTTGAGG GCCCGTTGTC CCTGATGTAA 9240
 10 ACATAATCGA TGCAATGTCA TCTAAATTAA ATGATGTATT TAATATGTTG GACGGCGACT 9300
 CTTTCGGCAC CACAGTTTCA TTCGATGTTT CATATTGGAT ACCCATTGTG TTGTCCAACA 9360
 AACTGTTTCG TGTAATATCC CTTCCAGCGA ATTCAATATC ATCCAGCGAT ACAATTTGAA 9420
 15 ACCCTCGTAA TTCCAGTGGC AAGGTACAAA AAATCAATTG TACATCGATT GACTTCATCT 9480
 GATTCGTCAT CTCATTAGGT GTCAACCTTG TATTAATCAT CGCAATTTCA ATATTTGCCA 9540
 20 ACCCAATGTC ATGTATTAAA ATGATCGATT GAATCGAATT ATCTATGTAT AGCCCAACAC 9600
 GAGATTGTTG ATAAGCCTTG AGTCTTTTAG CCAATAGACT CGCTTCACAG TATAAATTTT 9660
 GATAAGTATA AGATTCTTGA CCGTCTGTTA TCGCAATATG ATGTCCATTT TGTGTGCTT 9720
 25 GTTTATATAA CCAAAAGTCC ATGCGTTATT CCTCCAAAAT CATTIACATT ATAATTATAA 9780
 CGATTTTATG ACATTCTAGC AGTGGTTATG TTTAAAAATA TAAAAAAGTA GACGAATTGA 9840
 TGCATTGATA TGATTGTTAT AATGCTCAAT ACATATCGTT ATATCATTTC TCTACTATTA 9900
 30 TCAGTTATTT TTATTTAATT TTAGTGTCAT TCTGTCAATT TGATGTGGTG ATTTACCCAT 9960
 TGTGCCCACA TCATCTGCAA TGTCAATTGG TATACGGTTC ATGTCTTGTA ATGCACTTAA 10020
 ATGGAATACT TCATCATCTA AATTTTCAAT GAGATATACA TAATATGTTA CCTGTGCCTT 10080
 35 TTTATATTTT AACGTTTTC AAAAGTCCGG CTTGCAATTC AATACATTAT CCGGAATATA 10140
 TTCAATAAAT AAGTAACGTT TGCTGCCTAC TTTGTCTATG AAATATTTTG CAGTGCCTTT 10200
 40 TTCTATACCT CTTATATGTG CATAGTCTGC TGAAAAGTAA ATACTACCTA TTGTTTCATT 10260
 ATGTTGTGTG ATTTCAAATC GTTGGCCTAC TATTTTATTA TTTGTGCTAC nGGGGACTTA 10320

(2) INFORMATION FOR SEQ ID NO: 144:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1477 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

55

	GTGTGGATTG GATTTTAAAA TCACCCTCAT AAATACTGTC ATCAATATGA TAAGTTACAA	120
	TTTCACCTAT TATTAAATCA GCCCCATCTA ATACATCTCC AAGCAATATC ATTTGCGmTA	180
5	GTTTACATTC GAATCTCATT TTCGCATCTT TAATTCCTGG CGTCTTAATC GTTGTAGATG	240
	TTAAAAGTGA TAATCTGTGA CGACTCAACT CACTGTCACC ATATGCTAAC GGCGCTGCAG	300
10	TCTCATTAAT ATCTTGAACA TTATCTTCGT CTGTAATATG CACAACAAAG TCTCCAGTCC	360
	GTTCTATATT TAATGCAGTA TCTTTTCTCT TACCTCCTGC ACGTTGAACT GCAATAGCAA	420
	TCATTGGCGG ATGATTATTA ACAATATTAA AAAAGCTAAA TGGTGCTGCA TTTACTGATG	480
15	CATCTTGATT TAATGTTGTA ACAAAAGCTA TAGGTCGTGG AATAATTGAA CCAATTAATA	540
	ATTTATAGTT TTCTCTAGCA GTTAATGATT GTGCATCAAA CGTATACATA ATACCTACCT	600
	CTTTTCTAAG TATATCTAGG TATTTCTCCG ATTTTGTTA ATTTAAACAT CTATTCTCCT	660
20	CTGAAATCA CTGTATTTA TTTAGCAAAT CTTTGAAT ATGACACATA TGCATATCTT	720
	CTGGATATTT TTCTAAATGT TGCTGATGTT CTTCAGCACT TTTAATGTAG TTAGACAGCG	780
	GTAAGACTTC CACTGCAATT TGATCTCTGT CTTTACGTCG TTCAATGAAC TGACGCGCTT	840
25	CAATTAAGTG GTCATCTACA CAACTATATA AACCCGTTTC ATACTTTTGT CCAATATCAT	900
	TTCCTTGTTG ATTCACACTG TAAGGATCAA TGATTTCAA TAAATAATTC ATAATGTCTG	960
	TAATTGTTAA CATACGATCA TCGAAATGAA GTTTGACACA TTCAGCATAA CCATCATACG	1020
30	GACCGTCTAA TTAGAGCTT CTTCCATTG CTCTTCTGC TTCTGTATGT ATAATTCCAG	1080
	GTATTGTTGC AAAAAATGCT TCAACACCCC ATAAACATCC TCCTGCTACA TAAACAAC TG	1140
35	CCATATTTAC ACCTCATCAT CCTTTTTTAT ATTTTAAACA AGGTTATACC ATTTAATACC	1200
	GCCATGACAT GATTCTGATA CACCTTCATT ACGATACCCA TATTTTTCAT AAAATGAAAT	1260
	TAATGATTCT CGACATGTTA ACGTTACACC ATGTCGATGA TGATTCTTAG CAAGAGTTTC	1320
40	AAAATAGTTT AGTAAGCGAC CTGCAATACC CTGACCTGA TAATTTGGTG CTACAACAAG	1380
	ACCTAACACA CTAATATAGC CACCTTCACT ATTATTTGTG GAGACATTTT TAAATAAATC	1440
	ATCGCTAATG TAACGCTCTT TTATGACTGG ACCGTTG	1477

45 (2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3976 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55

	AGGTGATTAT CCTAAAAATG CTCATGAGGT CGCTATTAAT GATAAGTTAG CTGCAGACAA	60
	CATTAGAGTC GGGGATAGAT TACATTTTAA AAATAATTCA ACTAGTTATA GAGTTTCTGG	120
5	TATTTTAAAC GACACAATGT ATGCGCATAG TTCCATTGTG CTATTGAACG ATAACGGATT	180
	TAATGCATTG AATAAGGTTA ATACGGCATT TTATCCAGTG AAAAATTTAA CACAACAACA	240
	ACGTGATGAG CTTAATAAAA TAAATGACGT TCAAGTTGTG AGTGAAAAAG ATTTAACAGG	300
10	TAATATTGCG AGTTATCAAG CAGAGCAAGC ACCGTTAAAT ATGATGATTG TTAGTTTGT	360
	TGCTATTACA GCAATCGTTC TAAGTGCATT TTTCTATGTT ATGACGATTC AAAAAATATC	420
	ACAAATTGGC ATTTTGAAAG CAATTGGTAT TAAGACAAGA CATTTATTGA GTGCGTTAGT	480
15	TTTACAAATT TTAACACTAA CAATAATTGG GGTAGGTATT GCTGTGATCA TCATAGTAGG	540
	ACTATCATTT ATGATGCCGG TAACGATGCC TTTTACTTAA ACAACGCAA ATATTTTATT	600
20	AATGGTGGGG ATATTTATAT TAGTAGCGAT TTAGGTGCC TCACTATCAT TTATCAAATT	660
	ATTTAAAGTG GATCCTATCG AAGCAATTGG AGGTGCAGAA TAATGGCATT AGTCGTTGAA	720
	GATATCGTCA AAAATTTTCG AGAAGGTTTG TCTGAAACAA AAGTTTTTAA AGGTATTAAT	780
25	TTTGAAGTGG AACAAGGGGA ATTTGTCAAT TTAAATGGTG CCTCTGGTTC TGGGAAAACA	840
	ACATTGCTAA CGATATTAGG CGGATTGTGA AGTCAAACGA GTGGTACAGT GCTTTACAAT	900
	GATGCGCCAT TGTTTGATAA ACAGCATCGT CCTAGTGATT TACGATTGGA AGATATTGGT	960
30	TTTATTTTTC AATCTTCACA TTTAGTTCCT TATTTAAAAG TGATAGAGCA ATTGACACTC	1020
	GTAGGTCAAG AAGCGGGAAT GACCAAACAA CAAAGTTCAA CAAGAGCAAT ACAACTTTTG	1080
	AAAAATATTG GTTTAGAAGA TCGCTTGAAT GTATATCCGC ATCAGTTATC TGGCGGTGAA	1140
35	AAGCAACGTG TTGCGATTAT GAGAGCATTT ATGAATAATC CGAAAATCAT TTTAGCAGAT	1200
	GAGCCACAG CAAGTTTAGA TGCCGATAGA GCAACAAAAG TTGTTGAGAT GATACGTCAA	1260
40	CAAATTAAAG AACAACAAAT GATTGGTATT ATGATTACAC ACGATCGAAG ATTATTTGAA	1320
	TATGCAGATC GAGTGATTGA ATTAGAAGAT GGCAAAATAA CTGATTAGTG GCTTGTAAG	1380
	ACGCTAAATG TTAATGATTT AAGACATAGT AGTATAAAG TTAGATAACA GAATACGATT	1440
45	TGGGTTTACA AAAACAGGC TGGGACATTA AGTTCTTAGG CAATGTAAAA AAGCTGATTT	1500
	CTATTAATTA TTTGATAGAA ATCAGCTTTT TTGATATGTA TTTTATAATG TACAGCTCGT	1560
	TGCATTGATA TAGCTTGAAG TCACGTTTAA AACCATATCT ATCATTATGG TATGCATATC	1620
50	TTTTAAACC TATTCTTTTG TTATTAGGAC ATATAAATTC ATCATTAGT TCGTCATATT	1680
	TCCAATTTTG AGTGTTAAAA ATGTCACTTT TAACTTTCT AGTTTTATCT TTAATAAACA	1740
55		

	CACTATCATA ACATGCATCA GCTACAATAT ACTCCGGTAA ATAACCGAAG nTATTTTgAA	1860
	TCATTGTTAA AAATGGAATT AAAGTTCTAG TATCTGTTGG GTTTTGAAAT AGGTCATAGG	1920
5	ATAAAACAAA TTGAGAATTT GTCGCTATTT GTAAATTGTA TCCTGGCTTA AGTTGGCCAA	1980
	AGTGTCTTAT TTTTTTAAAG TATTTAAAAG TAAAATTACA TGTTAATACG TAGTATTAAT	2040
	GGCGAGACTC CTGAGGGAGC AGTGCCAGTC GAAGaCAGGG GCCCCAACAC AGAArcTGAC	2100
10	ATATAGTCAG CTTACAACAA TGTGCCGGTT GGGGTGGCTG AGACGGCACC CTAGGAAGGG	2160
	ACCCGTCATC AAAAATTCTA TTTATAGAAT TTTACAGTAA TGTGCCAGAT GGGCATAGCG	2220
	AAGcCATTCA ATACGAAGTA TTGTATAAAT AGAGAACAGC AGTAAGATAT TTTCTAATTG	2280
15	AAAATTATTT TACTGCTGTT TTTTTTAGGG ATTAATGTCC CAGACTCTTT AGTTTATTTA	2340
	TTTTCAATAT AACAAATTGC TAATCAAGGA TTAACGAATA TTTAAAGATA GTTTGACGCA	2400
20	ATATTAGAAA CAACCTATAA TAATAGTTTG TTTGTGGATT AACTATTATA AATAAAAGCG	2460
	GCGTAAAGAC ATATAAACCA ACTACTTGAA CAATATAACG TTAATAACAA TCTATACTGA	2520
	TACATTACGC CTAGATAATC TTTGATGAGC ACATGTAAGA AAAAGTGATA TGGTGTATGA	2580
25	CTTCCGACAC CATCGATAGA TAAACCTAAT TTTTGGGCTA GTCGTAAGGC GCGCAATACA	2640
	TGAAACTGAC TTGTtACACA AACAATTTTA ACTGCTTCAT GATACAAATT GTTGATGATT	2700
	TGTTTAGAAT ATAAAAAGTT TGTGTATGTA TTTATAGAGT GAGATTCCAT TAGTATATCT	2760
30	GTTTTATCAA CACCATGTGC AATCAAATAA CGTTGCATAG CTAAAGCTTC AGAAATTGGT	2820
	TCGTCTGGTC CTTGTCCGCC AGATACAATG ATCTTTGTTG CTGATGCTTG TTGTTGATAG	2880
	ATATCAAGTG CACGATCTAA ACGCGCTGCA AGCATTGGTG TGACAAATTC GGTAAAAATA	2940
35	CCAGCACCTA ACACAATTAT GATATCAACT TCTTTGTTGT ATGATCTATG TCTATATGAT	3000
	ACTGfCCAAA CGAGATAACA AATAAAGGTT AGTAACAGGG AAAGACATAA TATAGCTAAC	3060
40	CACATAGACA AACCTTTTAC AATAGGTGAC TGAATCGTAC TTATAAATAG AAGTGCTGAT	3120
	GTGTAGAGTA CAAATTTTATA TGAAAAAGAT AATAATTTTT TAATAAATAA GCGACTAGAA	3180
	GTATGAGAAA ATAAATATCT ATGTTTGAAT AGCATGATAA TACTGATTAT TATAAATGTT	3240
45	ACAAACATAG ACCAAGGGAA AGTATAGGTC ATGATGCTAT AGATGAGTGA CAAAAATATC	3300
	GATATGACAA CTAAGATGTA GCATGTTAAA TTTAACGTCA GAGTATAGTT GAAAATTAAC	3360
	GGACAAATAA CGATAAGTAT AAATATTAAT AATAAATTCA ATAACATACT GACACCTCGC	3420
50	TTATAATAAA TATTAAATAT AAATGTAGAT GATTTAATTT ATTAAAGCAA GGAGAAAGCA	3480
	GCAACATGTA AATCTTAATT TGTTATATTA TATATGGGTC AATATTTTTG TGTTTTTTAG	3540

55

TATGGTAAAA CATTTACAAG ACCATATTCA ATTTTITAGAG CAGTTTATAA ATAACGTAA 3660
 CGCATTAACT GCAAAAATGT TGAAAGATTT ACAAATGAA TATGAAATTT CATTAGAGCA 3720
 5 GTCTAACGTA TTAGGTATGT TAAATAAAGA ACCTTTGACA ATTAGTGAAA TCACGCAAAG 3780
 ACAAGGTGTA AATAAGGCCG CAGTAAGCCG ACGAATTAAA AAGTTAATCG ATGCTTAATT 3840
 AGTTAAGTTA GATAAACCAA ATTTAAATAT TGATCAACGT TTGAAATTCA TAACCTTAAC 3900
 10 TGACAAAGGT AgAGCATATT TGAAAGAACG TAATGCGATT ATGACAGATA TTGCGCAAGA 3960
 TATTACTAAT GATTTA 3976

(2) INFORMATION FOR SEQ ID NO: 146:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3346 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

25 GCTACCTAGG CATTTAAGAG ATCAAAAAAT GTATGAATAT GAACGTTATT TTTATGAGCA 60
 AGAACTTAAT GGCCTTGATG aAGGGGAAAT TTTAAAGAAG TTAAGAGACC CACAAGATGT 120
 TGCAGCTGAA ACAAAGCTA GAAGTGTTAT TGATTATGCT GAATCTAAAC CAACATTTGA 180
 30 AAATATTTCA AGAGCTGTTG CTGCTTCATT AAGTTTAGGC ATTCTATCTA TTTTGTGCAT 240
 CCTTATACCA GTATCTATAG TTGGATTATT TGTATTAGCA TTATTTTAA TATCACTTTT 300
 GCTGCTGTTT TGTCCAATTA TTTTATTAGC ATCAGCAATA TCCAGAGGAA TTGTGGACTC 360
 35 AATTAGTAAT GTATTTTTTG CCATATCATA TTCAGGATTA GGATTAGTAT TTATCATTGT 420
 CATATTTAAG ATTTTAGAAT ACATTTATCG TTTAATCTTA AAATATTTAC TTTGGTATAT 480
 TAAACTGTC AAAGGAAGCG TTAGAAAATG AAGAAATTCT TTTTATTGG GCTTTTAGTG 540
 40 TTTGTGTCT TTTTACAGC AGCAACCATT ATTTGGTTCA GCTATGATAA AAACAAATAT 600
 GGTACTAAAC AATATGATAA AACATTCAA gACGATGCTT TTGACAATGT ATCTATAAAT 660
 45 TTGGATAGTA CAGAACTTCG TATAAACGG GGAATCAAT TTAGAGTTAA ATATGATGGT 720
 GACAATGATA TATTAATTAA TATAGTAGAT AAGACGTTGA AGATTAGTGA TAAAAGGTCT 780
 AAGACAAGAG GATATGCAAT TGATATGAAT CCTTTTCATG AGAATAAGAA AACGTTAACG 840
 50 ATTGAAATGC CTGATAAAT GATTAAACGT TTAAATCTAT CATCTGGAGC AGGAAGTGTT 900
 AGAATCAGTG ATGTTGATTT AGAGAACACA AGTATTCAA GCATTAACGG TGAAGTAGTT 960

	AGTAAAAGTA ACATTAAAAA TAGCAATATT AAAGTTGTTA TTGGTACGCT ACAAATCGAC	1080
	AAGAGTCAAA TTAAACAATC CATATTTTTTA AACGATCATG GTGACATTGA ATTTAAAAAC	1140
5	ATGCCATCAA AAGTAGATGC AAAAGCTTCT ACTAAACAAG GAGATATTCG TTTTAAGTAT	1200
	GATAGTAAAC CTGAAGACAC TATACTAAAG CTAAATCCGG GAACGGGTGA TAGCGTAGTT	1260
	AAAAATAAAA CATTTACTAA TGGTAAAGTT GGGAAAAGCG ACAATGTTTT AGAATTTTAT	1320
10	ACGATTGATG GTAATATCAA AGTTGAATAA ATAAAGGATG TAAGCACCGA TATTAGGAAG	1380
	CATAATTTCT CTAATATCGG TGTTATTTAT TTGTTGGCAA AAGTTAAGTC GGTATCTATA	1440
15	TTGCCAGTAA AGTGAGTGAT ATTAAGGTCT TGACCATCTA ACCATGATTT GAAATCTATT	1500
	ATTTCTGGTG GCGCATTTTC TCCCAATGTA AAATATGCAG TTAATGTTTC AGGTTGATAC	1560
	ATTGATGTAT GGATGGTGCC AGACCAGCTT TTGAATAGTT TACTGTAAAT TTCATACTGA	1620
20	GGATTATTGA ATAACCTAAA TGCTGTAGTC ATATCTAAAT TATCATTAGT TTGTGAAATG	1680
	GTACGCGCCA GTCTTTCTTT AGATTCTTTT GTATAATTAC GATTTTCATG TGTTAATATT	1740
	TCAAAATGAT TTGTACATAT ATTATCATAA CGAACATCTA TTGATCTCGG TGTCACTTCA	1800
25	ACAATTGCAT GGTTCATGA TTTGTCCATC AGTATGTAGC TAAATGAGCT TCTGTGTGGT	1860
	ATTTCTTTCA ATAATTGGAT TGCTTCTGTT ACATTTCTGGC AATTTTCAAG AATTAGACGA	1920
	CCAATCATAT AACATACAAA ACCATTTGCT GGTTCCTTCC GGTGCATAAA GTTATAGCCC	1980
30	ATAGTTAATC CTGACTCATT CATACCATCC ATTCTTCCAG TTACCCTTGA TACAGGACCA	2040
	ATTTGAGCTA AACCGCTATC TGTAGGTTGA TAAAGTAAAGT AGCGACCATC ATAAGTTGCA	2100
35	GGGTGGTAAT CATAATTTCT AACCATGAAG TCTTTGCCTT GAAAGACCGT GCAaCCACTT	2160
	TCTTTTAAAT CGGTAAAACG ATAATGTCCA AAGTTTAAAA TAATTTGGCG TGTGGCATT	2220
	TTGAGTATAC TTTGTAGTCC CATTAAATCT TCCCATATTT GAGGTGCGTA TGTTTGGAAT	2280
40	ATTTGATAAG TTTCATTAC ATCTATATCG AAACGTGGGA CaChTTTTTT CCATTCTTTT	2340
	TCTCGATTTT TTAGAAGAGG TGTTTGTGTA AGCCATTAC CAGTTTAAAC ACCTAACTCG	2400
	AAATGTGAAC CTCTAAAAGT CATGATATCT GATGTCACTT GTTGCATATC ATCGGCCCCT	2460
45	TTCTTTTTAG TTGTAATATA TTGTAAATAA ATAGTAATCG TATGTATATT GAATGTCATG	2520
	TTAAATAAAG TTATATTTTA CTAAATGAAA TATAAAATTG TTTGAGGTGA TTTCTCGGTG	2580
	TATAAGACTT ATCAATCAGT TAAAACATAT TTTTATAGAT GGTGGGGATA TTGAGTTAAA	2640
50	AACTTAAAT CATCTTATCA TAAATATCAA TCTTAAGTTA GCATTCACGA TAATAGTCAT	2700
	TGTTAACATT AGCATATAAG GTCATGTCAC GTTGAAACAG AGGTTCTCG GCATTTTTGA	2760
55		

TTATTTAATG ATTATTCTAT ATATGATAGT ATAATGAAAT GTAGATAGGT ATTTAATTTA 2880
 ACAGAGGTGA AATTGAGATG TGGAATTTTA TTAAATGtGT GkTTAAATTC GTATTTAGCT 2940
 5 TAGTTGCTAT TACAACATTA GTTGCTGGTG TTGGTGTAGT AGCATTTGCT TATATCTTTA 3000
 AAAAAGATTT TGAAGATATT GAAAGAAAAA CTAAAGAAAT TATTTCTGAT ATTGAAAGTA 3060
 AAAATAACTA ATAACATTTA GAGGCTGGGA CATAAATCCC TAAAAACAG CAGTAAGATA 3120
 10 ATTTTCAATT AGAAAATATC TTAGTCTGT TCTCTATTTn ATcAmTACTt CGTATTGAAT 3180
 GGCTTCGCTT TCCTAGGGTG CCGTCTCAGC CTTGGTCTTC GACTGGCACT GCTCCCTCAG 3240
 GAGTCTCGCC ATTAATACTA CGTATTAACA TGTAATTTTA CTTTGGAAAT ACTTTTAAAA 3300
 15 AATAAGACAC TTTGGCCCAA CTTGGCACAT AAATGTAAAA TTCAAT 3346

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2375 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

GTTGAAGAAA GAAATATAAC AGTCAATTAT AATTATAACC TTGTTGAAAT CGACGGTGAC 60
 30 AAAAAAGTGG CTACATTCGA ACATATCAAA GCATACGATA GAAAAACAAT AAGTTATGAT 120
 ATGTTACATG TAACACCACC TATGGGTCCC TTAGATGTAG TAAAAGAAAG TACACTTTCA 180
 GATAGTGAGG GTTGGGTAGA TGTTAACCCA ACCACATTAC AGCATAAAAG CTA CTCTAAT 240
 35 GTATTTGCAC TTGGTGATGC TTCAAATGTA CCTACTTCAA AAACAGGCGC ACTATTcGTA 300
 AGCAAGCACC TATCGTCGCT AATAATTTAT TGCAAGTGAT GAATAATCAA ATGTTAACGC 360
 ATCATTATGA TGGTTATACT TCATGCCCTA TTGTTACTGG ATATAATAGG TTAATACTTG 420
 40 CAGAGTTTGA TTATAATAAA AATACTAAAG AAACAATGCC GTTTAATCAG GCCAAAGAAC 480
 GTaGAAGTAT GTATATATTT AAGAAAGATT TATTACCTAA AATGTATTGG TACGGCATGC 540
 TAAAAGGATT AATATAATAA AGTACAGAAA ACAATAAATT TTTAATGAAA AATCTTTTAC 600
 TATAAAAGAT TAAGTATTTA AATGACGTGT CAGTGTGTG TTTATATGTC GTGAATTTTT 660
 AGCTCTAAAT AGTATAAGAT TGAAAAAGTT GTTACTGTTT TAAATGATCA CGATGAAGTC 720
 50 ATTCAATAAG AATGATTATG AAAATAGAAA CAGCAGTAAG ATATTTTCTA ATTGAAAATC 780
 ATCTCACTGC TGTTTTTTTAA AGGTTTATAC CTCATCCTCT AAATTATTTA AAAATAATTA 840

	AGATATTCAA ACCACGTGTA CTCAAAATGA TAGCTTGGTA TGTACCTCCA ATAGTAATTT	960
	CAATAACTTT GTCTGTTGAA CACTAAGAGC AATTTTAATT TCATAATGTG TTGTAAACAT	1020
5	TTTTTTTGAT TGGAGTTTTT TTCTGAGTTA AACGATATCC TGATGTATTT TTAATTTTGC	1080
	ACCATTTCCA AAAGGATAAG TGACATAAGT AAAAAGGCAT CATCGGGAGT TATCCTATCA	1140
	GGAAAACCAA GATAATACCT AAGTAGAAAG TGTTCAATCC GTGTAAATTT GGGAAATATC	1200
10	ATCCATAAAC TTTATTACTC ATACTATAAT TCAATTTTAA CGTCTTCGTC CATTTGGGCT	1260
	TCAAATTCAT CGAGTAGTGC TCGTGCTTCT GCAATTGATT GTGTGTTTCA CAATTGATGT	1320
15	CGAAGTTCGC TAGCGCCTCT TATGCCACGC ACATAGATTT TAAAGAATCT ACGCAATCTC	1380
	TTGAATTGTC GTATTTTCATC TTTyTCATAT TTGTTAAACA ATGATArATG CAATCTCAAy	1440
	ArATCTAATA GTTCyTTGCT TGTGTGTTTCG CGTGGTTCCT TTTCAAAAGT GAATGGATTG	1500
20	TGGAAAATGC CTCTACCAAT CATGATGCCA TCAATACCAT ATTTTTCTGC AAGTTCAAGT	1560
	CCTGTTTTTC TATCGGGAAT ATCATCGTTA ATTGTTAACA ATGTGTTTGG TGCAATTTTCG	1620
	TCACGTAAAT TTTTAATAGC TTCGATTAAAT TCCCAATGTG CATCTACTTT ACTCATGCGT	1680
25	TTGATAAAAA CTAAATAAT ATTAATTCGG TCATCAGTGG CGTTAAATCT TTTATCATTT	1740
	TTAGTTATAG TTGATAAATT TATATTTATA AGCATATATG GATATTTTCA CAAAAATTTT	1800
	TATTTATATA AATCCGAACCT GCATACATAT TTGTTTAAAT AAGAGGTATT ATTTTTCGGG	1860
30	AAATTGCTGT CTGAGTTAAA AGGATTAGTT TTATAAAATG AGTTGAACTA TAGCCAAAAA	1920
	CGATTAAAT ACTGATAATC CATTTTTGtA TTATGTTAGG GACTTTTTTTA CTTAATTTTA	1980
35	ACCCTATTGG aGcMAATATA ATACTCCCTA TTATAAGGAA TAAGGCGTCA TATAAaGGGA	2040
	TATAACCTTG AATAAGTTTG ATGACAAAAG CACCAATTGA AGATATAAAA GCAATTACTA	2100
	TACTATTAGC GACTACAGTA TTCATTGGTA ATTTGAATAA AACCAATAAT ATAGGAATAA	2160
40	TAATGAAGGC ACCACCTGCA CCTACTATAC CTGAAATAAT ACCAATGAAA AGGCCAATGA	2220
	TAAC TAATAA ATATTTATTA AATGAAGACT TTTCGGAACT AGGTTtCACT TTAATAAACA	2280
	TTAATGTTAA TGCAAGTAAA GCAATAATGA TATATACCGT ATTTACAAAT GTAGCATCAA	2340
45	ATAAATTTGC TAGAAATGCA CCTAACATAC TCCCT	2375

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

- | | |
|----|-----------------------------|
| 50 | (A) LENGTH: 6115 base pairs |
| | (B) TYPE: nucleic acid |
| | (C) STRANDEDNESS: double |
| | (D) TOPOLOGY: linear |

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

	GAGGTTTCTA GACAAGCTTT TAATAACTTA CCAAACCTCAT TAAgrTGGTT gTgtTGGACT	60
5	GCctATTATC mAAGtATTAT GaGTTGTTTA ATATTAGtGC TAArACATAC GAAGAGTGGT	120
	TTAAACAATT TAGTAGTAAG AAAGCACAAT TCAGTATTAA TCTCACGGAT AAATGGATAA	180
	TTCAAATCGC ATATGGTAAA TTAATAATAA TGGCTAAAAA TAATGGCGAT ACATATTTTA	240
10	GAGTTCAAAC AATTAAAAAG CCAGGTAATT ATATTTTTTA CAAATATCGA TTAGAGATAC	300
	ATTCTAATTT ACCCAAATGT TTATTTCCGC TTACAGTGAG AACACGACAA AGTGGCGATA	360
	CATTTAAACT GAATGGGCGC GATGGTTATA AGAAAGTGAA TCGCCTGTTT ATAGATTGTA	420
15	AAGTGCCACA GTGGGTTCGG GATCAAATGC CAATCGTATT GGATAAACAA CAGCGCATT	480
	TTGCGGTAGG AGATTTATAT CAACAACAAA CAATAAAAAA ATGGATTATA ATTAGTAAAA	540
20	ATGGAGATGA ATAGCGTTAT GCATAATGAT TTGAAAGAAG TATTGTTAAC TGAAGAAGAT	600
	ATTCAAAATA TCTGTAAGGA ATTGGGAGCA CAATTAACAA AGGATTATCA AGGTAAACCA	660
	TTAGTATGCG TGGGTATCTT AAAAGGCTCA GCAATGTTTA TGTCAGATTT AATTAAACGA	720
25	ATTGATACCC ATTTATCAAT TGATTTTCATG GATGTTTCTA GTTATCACGG AGGCACTGAG	780
	TCAACTGGTG AAGTTCAAAT CATTAAAGAT TTAGGTTCTT CTATTGAAAA TAAAGACGTA	840
	TTAATTATTG AAGATATCTT AGAGACTGGT ACTACACTTA AGTCAATTAC TGAATTATTA	900
30	CAATCTAGAA AAGTTAATTC ATTAGAAATA GTTACTTTAT TAGATAAACC AAACCGTCGT	960
	AAAGCGGACA TTGAAGCTAA GTATGTAGGT AAAAAAATAC CAGATGaATT TGTTGTTGGT	1020
	TACGGTTTAG ATTATCGTGA ATTATACCGA AACTTACCAT ATATCGGTAC GTTAAACCT	1080
35	GAAGTGTATT CAAATTAATT TTtAATCAA TTtCAGTTAT TATTACTATG CGTTTGAGAA	1140
	ATAATAGTGT AGACTCAAAA ATATGAAAAA TGTATTTTCAT ATATATTTAA TTTTAGACAA	1200
	GACATATGTC TTGAAAAGTT GAAAAATATA GAGATTGATA AAAC TAATAC GGGTGTGAAT	1260
40	GACATTGATG TTAAGCTCAA TTAGTAGCTT ATAAAACATG TCATATGTTA CAATTTTTGT	1320
	TAGTTTTATT ATGGGAAGTA GGAGGAAATG ACGCATGCAG AAAGCTTTTC GCAATGTGCT	1380
45	AGTTATCGTA ATAATAGGCG TTATTATTTT TGGTCTATTT TCATATTTAA ACGGTAATGG	1440
	AAATATGCCG AAACAGCTTA CATATAATCA ATTTACTGAG AAGTTGGAAA AAGGTGACCT	1500
	TAAAACTTTA GAAATCCAAC CACAACAAAA TGTCTATATG GTAAGTGGTA AAACGAAAAA	1560
50	TGATGAAGAC TATTCATCAA CTATTTTATA TAACAACGAA AAAGAATTAC AAAAAATTAC	1620
	TGATGCTGCT AAAAAGCAAA ACGGTGTAAA ATTAACGATT AAAGAAGAAG AAAAACAAAG	1680
55		

	TTTCTTCCTA AGCCAAGCAC AAGGTGGCGG TAGTGGCGGT CGTATGATGA ACTTTGGTAA	1800
	ATCTAAAGCA AAAATGTACG ATAATAATAA ACGTCGTGTT CGTTTCTCTG ATGTAGCAGG	1860
5	GGCAGATGAA GAAAAACAAG AATTAATTGA AATTGTTGAT TTCTTGAAAG ATAATAAAAA	1920
	ATTCAAAGAA ATGGGATCTA GGATTCCTAA AGGTGTCCTA CTTGTTGGAC CTCCAGGTAC	1980
	TGGTAAACAA TTAATTGCTA GAGCGGTTGC AGGTGAAGCT GCGCACCATT TCTTCTCTAT	2040
10	TAGTGGTTCA GACTTTGTAG AGATGTTTGT TGGTGTGGT GCGAGCCGTG TTCGTGACTT	2100
	ATTGATAAT GCTAAGAAAA ACGCGCCTTG TATCATCTTT ATCGATGAGA TTGATGCTGT	2160
	TGGTCGTCAA CGTGGTGCAG GTGTTGGTGG CGGTCATGAT GAACGTGAAC AAACCCTAAA	2220
15	CCAATTATTA GTTGAAATGG ATGGTTTCGG TGAAAATGAA GGTATCATTG TGATAGCTGC	2280
	TACAAACCGT CCTGATATCC TTGACCCAGC CTTATTACGT CCAGGTCGTT TTGATAGACA	2340
20	AATTCAAGTT GGTGTCCTAG ATGTGAAAGG CCGTGAAGCA ATTCTTCATG TTCATGCTAA	2400
	AAACAAACCA CTTGATGAAA CGGTTGATTT AAAAGCAATT TCACAACGTA CACCTGGTTT	2460
	CTCAGGTGCT GATTTAGAGA ACTTATTAAA TGAAGCATCT TTAATTGCTG TACGTGAAGG	2520
25	TAAAAAGAAA ATTGACATGA GAGATATCGA AGAGGCAACG GATAGAGTTA TAGCCGACC	2580
	TGCTAAGAAA TCTCGAGTTA TTTCTAAGAA AGAACGTAAT ATTGTTGCTC ATCACGAAGC	2640
	TGGTCATACA ATTATCGGTA TGGTACTTGA TGAGGCAGAA GTAGTGCATA AAGTTACTAT	2700
30	TGTTCCACGT GGACAAGCAG GTGGTTATGC AATGATGCTA CCTAAACAAG ATCGTTTCTT	2760
	AATGACTGAA CAAGAGTTAT TAGATAAAAT CTGTGGTTTA CTTGGTGGAC GTGTATCAGA	2820
	AGATATTAAAC TTTAACGAAG TATCAACAGG TGCTTCAAAT GACTTCGAAC GTGCAACACA	2880
35	AATCGCACGC TCAATGGTTA CGCAATATGG TATGAGTAAA AAATTAGGAC CATTACAGTT	2940
	CGGTCATAGC AATGGTCAAG TATTCTTAGG TAAAGATATG CAAGGTGAGC CTAATTATTC	3000
40	AAGCCAAATC GCATATGAAA TTGATAAAGA AGTTCAACGA ATCGTTAAAG AACAAATACGA	3060
	ACGTTGTAAA CAAATTTTAT TAGAGCACAA AGAACATTA ATTTTAATTG CTGAAACATT	3120
	ATTAACAGAA GAAACATTAG TTGCTGAACA AATTCAATCA TTATTCTACG AAGGTAAATT	3180
45	ACCTGAAATT GATTATGATG CAGCTAAAGT TGTTAAAGAT GAAGATTCTG AATTTAATGA	3240
	TGGTAAATTC GGTAAATCTT ATGAAGAGAT TCGTAAAGAG CAATTAGAAG ATGGACAACG	3300
	TGACGAAAGT GAAGATCGTA AAGAAGAAAA AGATATTGCT GAGGATAAAA AAGAAGCTGA	3360
50	TAAATCTGAT GAAAAAGATG AACCAGCACA TCGACAAGCC CCAAATATCG AAAAACCTTA	3420
	CGATCCAAAT CACCCAGACA ATAAATAATC GATTATATTC AGTACCTCTT TCTATGATAA	3480

55

	AATTGTTATA GCAGAAAATA ATTGTAAAAC AAGTTACTTC ATTATTTAGA ATGATGGGTG	3600
	TAGAATAAGT ACAATTGTTG CATTTTATGA AGTAAAGTAA TTTTTTAAAT ATAGAGTAAT	3660
5	AGAGGAGATT GAAATAATGA CACACGATTA TATTGTAAAG GCATTAGCAT TTGATGGAGA	3720
	GATTAGGGCT TATGCTGCTT TGACAACTGA AACTGTTCAA GAAGCACAAA CGAGACATTA	3780
	TACATGGCCG ACAGCATCTG CTGCAATGGG AAGAACAATG caCAGCAACA GCTATGATGG	3840
10	GCGCAATGTT GAAAGGTGAT CAAAAATTAA CTGTCACTGT AGATGGCCAA GGACCTATTG	3900
	GACGAATTAT TGCCGATGCA AATGCTAAAG GCGAGGTGCG TGCTTATGTA GACCATCCAC	3960
15	AAACTCATTT TCCATTAAAT GAGCAAGGTA AACTTGATGT AAGACGAGCG GTAGGGACAA	4020
	ATGGATCTAT TATGGTTGTT AAAGACGTTG GAATGAAAGA CTATTTCTCT GGAGCAAGTC	4080
	CaATTGTTTC AGGAGAACTT GGTGAAGATT TTAATTATTA TTATGCTACA AGTGAACAAA	4140
20	CACCTTCATC GGTAGGTCTT GGTGTATTGG TAAATCCTGA TAATACGATT AAAGCAGCAG	4200
	GAGGATTTAT CATTCAAGTT ATGCCAGGTG CCAAAGATGA AACAATTTCA AAATTAGAAA	4260
	AAGCAATTAG TGAAATGACA CCAGTTTCTA AATTAATTGA ACAAGGATTA ACGCCAGAAG	4320
25	GATTACTAAA CGAAATCTTA GGTGAAGACC ATGTGCAAAT TTTAGAGAAA ATGCCTGTTC	4380
	AATTTGAATG TAATTGTAGT CATGAGAAAT TTTTAAATGC TATTAAAGGA TTGGGCGAGG	4440
	CTGAGATTCA AAATATGATT AAAGAAGATC ATGGTGCTGA AGCAGTATGT CATTTCTGTG	4500
30	GAAATAAATA TAAATATACT GAAGAAGAAT TAAACGTGTT GCTAGAAAAGT TTAGCGTAAT	4560
	TTAATTTAAA TCAATACGCT AAAATGTTTA TTTTACGCG TTTAGTGAAA TGTAGAACTA	4620
	AATAGTTGTA TAATCCTTAG TGATTTTGTT TGCTTTCTAG AATTTATTTG ATAAAATAAT	4680
35	TCTATATCCG ATAAATAAAC TAAGATTTCA ACAACTAACT AAAAAGGAGT GTTCTTAATG	4740
	GCAGAAAAAC CAGTAGATAA TATTACTCAA ATTATTGGCG GTACACCGGT AGTCAAATTG	4800
40	AGAAATGTAG TAGATGACAA TGCAGCAGAT GTTTATGTAA AATTGGAATA TCAAAATCCA	4860
	GGTGGTTCTG TAAAGGATAG AATTGCTTTA GCAATGATTG AAAAAGCAGA GCGAGAAGGC	4920
	AAAATTAAAC CTGGCGATAC AATTGTAGAA CCAACAAGTG GTAATACAGG TATCGGTTTA	4980
45	GCATTTGTAT GTGCTGCTAA AGGATATAAA GCAGTATTTA CTATGCCCGA AACAATGAGC	5040
	CAAGAGCGTC GTAATTTATT AAAAGCATAC GGTGCGGAAT TAGTTTTAAC GCCTGGATCA	5100
	GAAGCGATGA AAGGTGCAAT TAAAAAGCT AAAGAATTGA AAGAAGAACA TGGTTACTTC	5160
50	GAGCCACAAC AATTTGAAAA CCCTGCGAAC CCTGAAGTTC ATGAGTTAAC TACAGGTCCT	5220
	GAGTTATTAC AACAATTTGA AGGGAAAACT ATCGATGCGT TCCTAGCTGG TGTTGGTACT	5280

55

GTTGCTATAG AGCCTGAGGC TTCTCCAGTA TTGAGCGGTG GTGAGCCAGG TCCACATAAA 5400
 TTACAAGGTT TAGGTGCTGG ATTTATTCCA GGCACTTTGA ATACAGAAAT CTATGACAGT 5460
 5 ATTATTAAAG TAGGAAATGA TACAGCGATG GAAATGTCTC GTCGAGTTGC TAAAGAGGAA 5520
 GGTATTTTAG CAGGTATTTT ATCAGGTGCT GCGATTTATG CTGCCATTCA AAAAGCAAAA 5580
 GAATTAGGAA AAGGTAAAC AGTAGTAACA GTATTGCCGA GTAATGGTGA ACGCTACTTA 5640
 10 TCAACACCTT TATATTCATT CGATGACTAA TTAATGTCAT TAAAAGAGT GAGTTATCTT 5700
 TTTGAGATAA CTTGCTCTTT TTTTCTACCA TGTATATTTT TAAAAATATG AGCGTTAAAT 5760
 TAAACATTTT TCTGATAAAA ATATCCAGTG AATGATAAGA TAATAAACGT ACATACTAAT 5820
 15 AACTAGTAAA TAGCAGGAGT AAATTTTATT AGAGTTAAAC AATACATAAT TAAAGGGTGG 5880
 TTAACATGAC TAAACAAAA ATTATGGGcA TATTAAACGT CACACCTGAT TcATTCTcAG 5940
 20 ATGGTGGAAG ATTAAATAAT GTTGAATCAG CTATAAATAG aGTGAAAGCC ATGATAGATG 6000
 AAGGTGCTGA CATTATAGAT GTTGGAGGTG TTTCAACGAG ACCCGGTCAT GAAATGGTTT 6060
 CATTAGAAGA TGAGATGAAC AGAGTATTAC CTGTTGTTGA AGCTATTGTC GGTTT 6115

25 (2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10401 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TAGATACTGG GnTAAaCaTc AAAAATAtyT GcTtATTCaC GTGTTTAcGc TCCcTCAAAC 60
 GCAACGTTAA TTGCGTGTA TCATTTAGTG TGAATTcAGA CGCTTCTTCC ATGACTATGT 120
 40 CTGATATGCC TTTTATCGAC TTTATTTTCT CTGGGTTATC TAATCCTTTA AACAAAAAAA 180
 CTGCGCCGTT TGGCAATTCA ACTTTGTTAT CAGTCTTATT CCAAAGGCAC ATGTCCCAAA 240
 TACCAAAGTT TATCAAACAA TCTTTAACAT CTTCGAACAA ACTATCTTTA ATTGTTGATT 300
 45 GTACTTTTCT AAGCCACAGT ATACGCCTAG GATATTTCCA ATCTTGCAAT GCTTTGAGTA 360
 CAACTTTTGT TATAACGCCG TGAGACTTAC CGCTCGAACC TCCACCGTAA TGkACTTCAG 420
 TGAAGTtATC GTAATTGGTT AGTATTTCTGA ATATGTTTCT ATTGAAAACA TTAGACGGTT 480
 50 TGTTAAAGTT TAATTTAACT TTCGTCATCG TACTCACCAA TATTAATCTC AATATTCTTC 540
 TGAGTAATTT CTTTTTATC GATATACGCA CCATGTACTT TTAGTATGTG GTCAATAGAT 600

55

	TTTAAATGGT CATATTTCTT ACTGTAAGCC TCTTGAGGTT CTCCTCTAGC AATAGAAGCA	720
	GATAACGCTA AAGCTTCTGT AATACTCATT AAACGCTCTT CTTGTATCTG TTCTAATCGT	780
5	TCTTTAATAT ATTCCGAAAC ATTAACATTT CTTAACAATC GACTTGCTAA AGACTCTGCT	840
	GTTTTCTTAC TATAACCTGC TGTAAATGCT GCTTTTTTAC CATTACATCC ATTCATTATA	900
	TATTCATCTG CGAATCTCTT TTGTTTTTCG TTCATTTTCAT TTACCACCAA CTCTCGCGCT	960
10	ATACGCTTTT TAAAATTAAA AAAGGATTGG CTATAATCAG CCAACCCACA TAGATCCTTT	1020
	ATTCCTAATT GCGATAAGGG AAACGCAGTA CGATAGTCAA TATCCTACAC TATCATAATA	1080
	TCTCATTTAA GGTATCAAAA ACTGCCACTT TACTGCCAAT TTCAGTCTTC CCCTAACTCT	1140
15	TCCGCCAATC TAGATATGAT TTTTCTTTTG ATTCTATGAG CAGTTCTATC AGAAATGTGT	1200
	ATGTCAACAC AAACCTTTCAC TAATTCCTTT TTATTAATAA AATACTCTTG AATGAATTCG	1260
20	CGTTCCTTCC TGCTTGATGT GTTGATTATA CGTTCAATAG CGCTCTTAAA CTCAAGGATT	1320
	TTACCTCTTC GTATACTACA AAGATAATTA GTTACTGCCA TTTCTGTTTT CGATGTATTA	1380
	GACGGTACAA ACTCCCCGCC TATATTTGTA TCTGTTGGAA TCCACGGTGT CATTATTTCA	1440
25	CTTCTTAAAT CTTCAAGTTG TTTATGATAA TTAGGATAAT CACACAATC ATCTTCTAAC	1500
	TTTCGAACTG TTGATAATTT TAATCCGTAT TTCTTTTTAG TCATGAATAC CCTCCGTACA	1560
	AATATGTTTA ATCTTCAAAG TGTCTCAATC TACTTCTTAA TATCTCTATC TCTCGCTCTT	1620
30	TAACCTTTAC ATCACCTTTT AACTGTTCCG CTTGTAACAT CACACCAAAC AATAAGATGA	1680
	CTAGTAATAT AATTGCTATG ATTAACCACA TCATCTACTC CGACACCTCC GCCCTCATCA	1740
	AATCAGACTG ATCACTCAAC TTTGCGAAGT CACTTGGCGC CTCTACATCA TCATTAGCCG	1800
35	TCATCATAAT ATATACTTGC TCAGTTACAT ACTTACCTAA CTCATACATC GCTAGTAAGA	1860
	ATAA ⁵ AGTCT CAAAATTTCT TTAACCACCA CTAAACACCC CATGTTAATT TATCGATAAT	1920
40	TTGTATAGCT TGTTTTAATG CGTCTCTTTT TTCTTTGATA TCTCTATTAT CGCCATCTTC	1980
	ATCAGCTGAC ATTAACTCAC TGTCATATTC ATATAATAGT TCTGATATTT CATTACTAGC	2040
	TACTACTAAT AAGTTTTTCAT CTACATCAAT CGTTACCGTT TTCTTTGGCA TCTCCATCTC	2100
45	TCCTTATCTT AACTTGTGCC TCGTATTTGC GCTCAGCTTC TTCTTTACTC TCTGCCTCAA	2160
	CAACTGTAAA CGTCTGATTA TCTCTAGCAG TAGTAAAATG TTCATGTGGT TGTCTGTG	2220
	AATCTTTGAA TGTTGTGACT AAGTATTGCG TCACCTCTTA TCACTCCTTT GAATGATTCT	2280
50	AAGTTTTTCT ACGAATAAAA GTATTAGTAC AACACTCAAT GTAGCCAACA TATTTTTTTG	2340
	CTTTGCAAAA TCTACTATAA CGATTAAGAC TAATAACATT CCAATTCTGC ATGTAAATAA	2400

55

	TACAAGTATT	GGAACATAATG	TAATGATGTA	ACTCACTTCC	CCAAAACCTC	CTTGACTCGA	2520
	TCTAAGATGT	CTTTACACTC	CGCTACTTCC	GAAGCCTTTT	TCTCCACGTT	CTGAAACACT	2580
5	TTCGAATTCC	TCCACTTGCT	TTAGTTCAGG	TGTCCATATA	GGCACGATAA	CCAATTGAGC	2640
	TAGTTTGCT	CCTTCGTTGA	TTTGATAAGT	TCCGTATTGT	CTTATGGCGT	CACTCAAATC	2700
	GATTTCTCCT	TTAATATCAA	AAACACCTGG	TGTGATATAA	CCATTGATG	CAATAGCGTC	2760
10	ATTCTTGATA	TTAATCCCTA	AATTGCCGTG	ATATCCCGCG	TCTATCTTGC	CTGTTTCAAT	2820
	CACTAAATGC	GTTTTACTAC	TTACACCACT	ACGGCTAGTT	AATAGTCCGA	CATAGCCCTC	2880
	TGGTATGCTT	ACAGCTACAT	CTGTTTTAAT	CACTGCCTTT	TCTTGTGGCT	CAAGTACGAC	2940
15	AGTTTCAGCT	GAGAATATGT	CATAACCTGC	ATCCGTCTTA	TGATTTGCTT	CGGGCATTCT	3000
	AGCATTTTCT	GATAATAGCC	TTACTTGTA	TGTGTTAGTC	ATTTTCCTGC	TCCTCCCTAG	3060
20	CTGTAGCAAA	CGCTATTCTC	AATTTCAATC	TTTCAACAAT	ATGAATTAGT	GCGGTATTGA	3120
	GGAATATTTT	AAATTCTTCA	ATGTTCTCAT	CTATAAAATC	AAGTATTTCT	TCCTCTTGTT	3180
	CACTGTCAAA	CTCGCTTAGT	ACATCCCCAA	TATTTATGTC	GCTTTTGCTC	GTTTCTAATA	3240
25	CTCTTTTGAT	TATTTCTGAA	TTACTTTTAT	TACTCATTTT	CCTTGTTTCT	CCTCATATTT	3300
	ATAGACAACT	TGACCTGCCA	TAATCCCTAC	TGCTTCATCA	AGTTCAATAC	CTTCTTTAAC	3360
	TGAATGTTGA	ATAGCATTTG	TCATTCCCTC	AAGTATTTCA	TCAAACGCTT	GTGCTCTCTT	3420
30	ATACACGTCC	TCAATCTCTT	TTAGTAATCC	CTCTGTGTCA	TTACCGTTAT	ACGCACTAGC	3480
	ACTGATCACT	GATTGTTCAA	TTGTTTCGCG	GTTATTCATC	ATTTCCATCT	CCTCTAAAAT	3540
	AAAGTTAGTT	GCTTCTGCTC	CTCGTATTCC	AAACCATGTT	GCTTTATATA	TGTTTCGAGC	3600
35	TCTTCCGCTG	TATCAAATGT	CTTTTTACG	CCTTGCCAAC	CTGGCACGAT	ATGCCCATGa	3660
	AAGTAATAAG	TGCCGTTTAC	TACATGGATA	TGTGCCACTC	GTTTCGTTATC	CTGATACAGA	3720
	TATCTCTTAG	ATCCGAAAAA	TTGGTTTAA	TATTTCTTAC	ATGCGCTATC	GGTTTTAGGC	3780
40	ATTTATGCTT	CCTGCCATTT	CTTAAACATT	TGGTTATAAG	TAGTATCAAA	CCAGTACGGA	3840
	TCACGTGAAT	GTTTTTGAGG	CACATTAAAC	AAATGTGGCT	TCTTCTTACG	TAGTTCAGCC	3900
45	TCTTTACGTC	GTTGCCTAGC	CATTTACGCG	TCTTTGCTCT	CTCGCTCCAT	GATTTTGGAT	3960
	AACACAATTT	CTTTATACTC	AGCTAAGCGC	ATACCATAAG	GTGCATGTAA	GGCTTCTAAC	4020
	AACGCCCAGC	CACCTCGTAC	TCTTTTTGCA	ACCATTCTTG	GAGTTAAACC	GTTCTTTTTT	4080
50	ATCAATTCAT	TTTCATGTTC	GGTAAATTTA	TATGGTTTAc	CGTTAATCTT	TACGATACTC	4140
	ATTTATTCCA	CCTCTATACA	TTTACTTTTT	TTAATCCAAT	CCTCTAATTT	GTGCGTGTG	4200

55

	ACATTTAAGT TAACCATCTC AGCTTTTCCG TTTTATATC CACTAATAGT TGATCTTGAT	4320
	ACGCCAGTTT CATTTGTGCAA ATCTTGGACA CTTACGTTAT CTCTAGCCAT GATTACCCCTT	4380
5	AAATTAGTTG CGAATACTtC GTTCAACTTC ATTTATTCCA CCTCTATATA TGCATGTCTT	4440
	ATTGTTATGT TGTCACTACT TAGTAATTCG TCCGGATTGT CATCTAAGCG CTTTGCCAGC	4500
	GTATCTTTTT CTTTATCCAC ATCATCGTAA TGCTGATATT CAACTTCTGT AGGTATTCTT	4560
10	ATATCAATCG TTGCGTTTAT ATATGCTTGT TGTTGCATTA GATCACTTCA TTTCTCTTTT	4620
	TCTTTTACGT CTGACTTTCA CTAAGTCCTC ATATACCATC CATTCTTGAC CTGTGTATTT	4680
	AGGCGCTTTA CATATCCACG TTAAATTCAC ATCTCTATAC TGATATCTGA ATATCTTCGC	4740
15	TTTGATGTTG GCAACTTCAG TCGCCTTACC TTAAACGTCT ATAACCTCAA CCAGTTTCCC	4800
	TTCTTCCAC AAAGAGAAAT CGGCTATATA CGTAATCGGT CTTTGTTC CGAATTTAGG	4860
	TTGTAATTCA AATTTTCGGT GTATTTTCGAT ACGATCATAG TTAGTGCCAT TCATATTACT	4920
20	TTCTAAATAT TGGTAATATT CGCACTCTAC TTTGCTATCA AATACAATTC CTTTGTACTC	4980
	AACTTTCTTA GCATTGTATT TACTCATTGT GCCACCTCTA AATATCAAAT ATCGTTGCTT	5040
25	GCAATCCTAG CTCTTGCTCA TATAGAAGCC CGTGAGCGCC TTTGAATCGT TTTAGGTCAC	5100
	TATCAGTCAT AATTTTCTTT TCGTCGCTGA AATGGGCTCC TGTGAGCGAA TAAACTTCAT	5160
	TTACGTTGTC TTTATACTTG ATGACCTTAA TATCTTCCGT GCCATCTTCT CGGTATAAGT	5220
30	AATATTTTTT TTTCCGGCATT TTTTAACACT CCTTAATGTG TGTTCCTTC CAGTTGATTT	5280
	CATTCATGAT TTTCTTTTCA ACTCTGTCGT AATCATCGAA AGGCGATAAC TCGTTATTGT	5340
	CCAACAATCT ATTGACCGCC CAACCACTCT CGATATATAC ATTTGCTACA ATCGGGTCGC	5400
35	TTTGCTTTGT CTCTTCATAC ATCGATTTCATAAAGCTTTT GAATTGCATT ATATTCATGT	5460
	GAAAACCTC TGAGTCTTCT TGTAATACTC AAATTCAAAT ATTCCGGTTT CGCCGTCTTT	5520
	GTTTTTGGCT ATGTTACATT CAACAATAGA TTTGCCAGTG ATACTGTCAT CTTGTCACG	5580
40	GTTATAATAA TCATCACGGT AAAGTAGCAT CGCTAACTC GCATCTGCTT CTATTCCGCC	5640
	TGATTCTTTC ATGTCCGATA GCATTGGTCT TTTATCCTGT CTAGACTCGA CACCACGATT	5700
	CAGTTGTGAA AGTAGTACGA TGATTGCGCC TGTCTCGTTA GCGATTATCT TTAAGTCACG	5760
45	TGATATCTTT TCTACTGCTA CACGTCTATC AACTTTCGCA TCAGTATCCA TCAGTTGAAG	5820
	ATAATCTATA AAAATAACTT GTTGCCTGTC TGAATGCCTC ATTGtTGCGC TCGCACATCT	5880
50	TGCGGTGTGA TATTACTTTT ATCAGAAATA TCGATGCCTA ATTTTCATGAT TTTATCCATC	5940
	GCATTCGTTA ACTTTGTATA GTCATCCGGC GTTAAGTTCC TGATTTCTTT TATCTTTGTT	6000

55

	AGACTAAAGA AAGATGTTTT GTATCCATTT TGTGCTATGT TCAGCATCAT GTTTAATGCA	6120
	AAACCTGTCT TACCCACTGA GGGACGCGCT GCGATGACGA TTAATTGTGA TGGTTCTAAT	6180
5	CCCCCTATTT TGTAATCCAT TAGCTTGTA CCCGTCTTAA TTGCTTCTT AGGGCTATCG	6240
	CTGTATAACT CTTCGACAAA CTCCTCAACA AACTTCTTGG TTCCATCTTC TTTTTGTGA	6300
	GTAATTGTTT TTAAATCCTT GAGTTCATCA ATCAAGTTGT TAAAGTTTGT GTTCGTAGGT	6360
10	TGTTGTTTGA ACTCAGTTAC CAATTCGTGA GCTTTGTTGA GCTGATAACT TTCCAATAAT	6420
	TCTTGTTGAT AACGTTCAAA GAAGCCATAT CCAATGAAAT CGGAGTTGTA AAGTTTAGTT	6480
	ATAGTATCTG CATCTAAAAA TTCTTTATCT TTAGTTGCTT TTAAATAGAT TTCTTGATGA	6540
15	TCTATCTTTC CGACGTCCAT TACATAATTG AAAAAGGTTT TAAACTTTTC GTTCGTAAAC	6600
	ATGTAATCTT TAACTCTTAT CTTTCTAAT ACGTCCGTT GTTTAAGTAG CGTAGCGATT	6660
20	ATTGTACTTT CAATTCGAA TTGTCCGTAA TTCATTGTT TTCGCCCCA AATTCTGCCA	6720
	ACTTATTCAT GAACCTATCT AGCGCTATTT TTCTTTGTCT GACATATTCG GGGTCATTCT	6780
	GCATTTTCCA TTGGTGTGTA GCGGTTTCGT TATCTACTGG CTCGATAGAT ACTTTTTTAG	6840
25	GTTCCTTACG CATGATTGCT GGTAAGTTAG GCGGGTACGG GTTGTTACTG TTGATATAAA	6900
	CATCTACCGC TTTTACAGTT GGTGATAAT CTCCATTTTG ACTTAATACA TCAATCCACA	6960
	TTTCTAACTT CGGTTTATCA AAATCAATGT TGTATACGTA CCTAACTTTT TTAATAATTT	7020
30	CTAATGCTTG TGTTTTGCTC ATCGGCATTA GTCATCACTC AATTCTTTTT CCATTTGTGC	7080
	AATGACATCA TCAGTAGTAT TTTTCTAGG TGCTATTTTA TTTTCTGCAT CTTCTTTTGT	7140
	TTTGACATTC TCTTAGCCC AGTTGTTTAA AACTTTAATT AAATAGCCAC CATGCGCACT	7200
35	TTTGCTTTTA GTGTACTCAA CACCTACTTT TACAACTTCA AAAGCGTTTG TACCTATATC	7260
	ATCAATAGCA AACCCTAATT GTTCCATTG ATTAGGTGTT AACTTATCAT CCAAATTGTC	7320
40	AATTATATAT TTATTGAAG ATGAGAAGAC GGCTTCTCTT TCTTCTCTT TATTCTTATA	7380
	TTCTTCTTCT TTTTCTTCT CTCTTCTTC TTCTTCTTCT GTATCGTTAC GTAACGTTAC	7440
	GGTAACGTTA CGTTTTGCTT CTAGTAACCT TTTCTGTTTC TCACGATAGC GTTGTTGTCG	7500
45	CAATTTATTT TTTTCTTTAT GCTTAGCTTT GCTATCTAAG CTTTGATGCT TCTCCAGTT	7560
	TGTCACTTTT ATGACACCAT TAACTTTTTT AATCATGCCC AATGTCTCAA AAGTTTGAAT	7620
	TGCTAACCTT ATTGAGTTAA TAGGTCTATT AAATTCATTT GCTAACATTT CTTCGTTGTA	7680
50	CGGCAAGTTT TCGGATAGCA TAATATAACC TTGTTTCATTG TACTTTCTCTG ATAAAGTTAG	7740
	TAACTTAACC CAAATAGTTA TGATCGTATC TCTTTCGGGT AAAGCTTCGA TATATTGAT	7800

55

	CTCCTTTCAG CATTTTGTG AGCCTCTCAT CAACTTTTAT CCACGAGTCA TGCAAGTGAT	7920
	ATTTATCATC AAACGACTTA ACGCCAATTG CGTGCTGTTC ATTATGATGT TGTCTACACA	7980
5	GTGCTAACAC ATGTTTGTG TAGTGATTCA TTTTGTCTTCT GTTCATGCCT CTGCCGACTG	8040
	CTTCATAATG TGCCAGGTCT GCGTGAGGCT TTCCGCATAT TACACAGTTG CGGTTGATTG	8100
10	TAGCCCAATA TAATAACGCT TTATCTTCGC TTAACAACCTT ACTCGTTTCT ACACTCATAG	8160
	GTATTTGATG ATGAAACATA AACGCTATAA TCAGTTCTAT TAACTCCCTT GCAACTTTCA	8220
	TAGAACAGTC GCGCAGACTG ATTTCTTCAT AACCTTTCAT AATTTCCAAT TCTGTTTGTA	8280
15	ATAATTTTCT AGTTGATTCT ACTGGTTCGC CCCAGTGAAG TTCTATATCT CTACACATTG	8340
	CGAATATTTT TTTGCGTTGT TCTATAGATA GTTTTTTATT GTCCGGAACC TCTACTTCTG	8400
	CTTTTAGTGG ATATCCGTTT TCTAGTAAGT CAATGTGACT TTGTTCAAGT TCAACACCAG	8460
20	TAGCAACGAC GGAATAAGTA CCGTCATTGT CTTTCTGGTA TCTTGTAATG TATTGCATTT	8520
	AAACCACGTC CTAGAACGGT AAATCATCAT CATTGATTTT TATTGGACCA TTAGCATTAG	8580
	CGAATGGGTT TGATTGTTGA CTCATTGGCG TCTGTTTCCC ATTTGCTTGC TGTTCTTTTT	8640
25	GTTTCATCTC ATCAGTTTTA GGTTCTGGTT TATTAAC TAC TTCATCGTCT TTATTCCAAA	8700
	CTTTTACATA TGAGAGTCTT ACAAATACT TGCCTTGTTT CTCGTTAAAT TTATTTTAA	8760
	GTACAATAGT TCCGATTTTG TTAATTAATT GATCTGTGTC AAAAGTTAAA TCTGGTAAGT	8820
30	TCAATTTAAT TCCTAATCTA CTAAGTAACT CGATATATTG TTTTCTTGA TAATCTTGTT	8880
	GGAATGGTGG GACGAATTGG TTGTGTTTGT ATTGTTTACC TTCGTTGTTT TCAAAAACAA	8940
	TCGTGAAGTA TCTGTTTTCT CTGTCGTTAA ACTCGACATT TGCAACTTTT ACTGTAAATT	9000
35	CTCCAGCTCC TAAAAAGTCC CCACCTTTCA TGAATGCCTC TTGATTAGTT TCTTGAATGT	9060
	ATTGTTTCT ACCAGTGATT TTCATAATTT TTATACCGTC CTTTAAATTA ATTTTAAATT	9120
40	ACCATTTCTA ATTGCTTGTA CAACATCGTT AATACTTGGA TTAATGAAAC GTTTGTTGTT	9180
	AATTTTGATG TTGCTTGAGT GTCTTATCTT TGTCTCGAAT AAATTTGATG GTTCAGCGTT	9240
	AAGTACATAT TGATAAGTTT TTTGCCGTC TTGCTCATGT TCTTCTATTG TCATTCTTGC	9300
45	TAACACGTCA GATTGACTGA TGA CTGCTTT TTTTATTTGG TCTTGTGCCT CTATCGTGAT	9360
	TGTTGGATTG ATAGTACTTC CCTCATCATC TTTGCTTTG TTAATGCCCT CGTGTCGCT	9420
	TATAGCAAGA TGAAATTGAT AATGTTCTTG TAATTTAGAA ATATAACGAT AAATACTTAC	9480
50	AATGCGTGTA GCACACTCGC CCCAATCATT AAATGTCGGT TTCTTTGATT TACCGTCCAT	9540
	GATGTCGTCC ATAGTGATAT CACGTAACCT TTGGATTGTT TCAATCACTA CAACATCAAT	9600

55

AAAATGCTTA TAATTCTTAA TCTGCACAAC TGCCCCATCT TCTGTTACCG TTGTTCCGTC 9720
 CTCATTTATA TCTAGTACTA AGGCATTGTT ATCTTTTGTT AAAAACGTAG TTTTACCAGT 9780
 5 ACCGAACTTG CCGTATATCG CAAATTTATA AAACCTTGTT GCATTTTGTT TGCTGATGTC 9840
 TTTTACACCT AGTTGCGTTA AAATATCGAC ATCTTGATTA GTTTTTTCAG TCATCTATTC 9900
 TCCCACCTTT ACCGTGTATG ACGTTGGTTT CTCCACAATG CTAGCACCTT CTAAAACTTC 9960
 10 GCCGTTTGCG TCAATCAATG TGCCGTTTTT AGTTACATTG AAATCTTTCT TAATGTCTGA 10020
 TTGGCTAAGT TTTTtagTTA CTTTACATA GTTGTCAAAA CCTCGTTGCT CAAGTTGnT 10080
 AATGACTTCT TGCTCATTGC TAACTTGAAT GACTTTTGAA CCTTTTCTGG CTGTCACCTT 10140
 15 TCCGTAAGtG TATTCAACTT GAATTTGCTA TCTTGTCTT TTTGTATTCT GTAATATTCA 10200
 ATTACAAGGC TTTGTAAATA TTCTTTGCCA CTCTGTAATT TTTCTACTTC TTTATCTTTC 10260
 20 CATTCGTTTA TGC GTTCAAT TTCTTTATTT GCTAAATCGT TGATTTCATT CTCTTAGTT 10320
 GTGATTGCAT CCAGTTTCTn AAAAACCCAG TTAGCACTGT CTAGATCAGT nACTTTGAAT 10380
 CGGTCGTCTT GTTCGAATGT n 10401

25 (2) INFORMATION FOR SEQ ID NO: 150:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2989 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

35 TTTCTCTCTA TTATTCTCGA TCGTAGATA ATTGTTTAAA TTTAAGTTTA TAGTAATGTT 60
 GAGTTTATAA TTTTATATAT CTAAAAACAG GTGTTGTATA TATAATCATT CATCTAGTTA 120
 TACTTACTTT AAAAATAATA TAATTTTCATG CGATGCAATT CATTGATGGA TGTTTTTAAT 180
 40 CTTAATCAAA TCCAAATAAA GCATATATTT TTAAATTCAC TTTCTTTTGA ATCGATTTTT 240
 ATCTCTTGnA TTAAACTTTT CCATTGTTTC ATTAAAGCTC TCTGTCATAT CTATTCCCAT 300
 TGAATTCGCT AAACATAACA ACACAAATAA ATTATCACCT AATTCGCTT TAATCGTATT 360
 TGCTTCTCTT GAATCTTTCT TCTTTTTTTC ACCATAGGTA TGATTATTTT CACGTGCAAG 420
 TTCGCCCACT TCTTCAGTCA ATCTAGCTAA GTTAGCTAAT GGTGAAAAAT ATCCTGTTTT 480
 50 AAATTGTCCA ATATATTCAT CAACCTCAGG TTGCATTTCT ACCATTGATT TCATTTCTAC 540
 GTTCTCCTTA TATTGCATTT CTAATATAGT ATATATCAAT TTGAAGTCTC ATGCATGTTT 600

55

	AATTCAGTTT ATATAAATGT AATGCATTCC TAACTAAATT AAATCAATTG AAATTGGGAT	720
	TATAACTTTA TGATACGTAC CACTACAATA AAATAATATA GTGAATAATC TACCATTAGA	780
5	AAAATAAGCA CAAAAAACT AGCAACCACA CAAAAATGTG ATTAGCTAGT TAATAAGTGT	840
	CTAATTTAAG TTAATTGTTA ATCTATAAGA TTAATCACTT GAACGCGCAA TCAAAATAAT	900
	ACGTACAAGC TCTGCTACAG CGACTGCAGT TGCTGCAACA TAAGTCATTG CTGCTGCAGA	960
10	TAATACTTTA CGCGCATGCT TGTATTCTTT TTCATTTACA ATGTTCAATG CCGTAATTTG	1020
	TTTCATCGCT CTTGAACTCG CATCAAATC AACTGGTAAC GTAACAATTG AGAATAATAC	1080
	CGCTAATGAC ATTAAACCAG CACCAATCCA TAAAGCAGTT GAACCaAATG CACTACCTAT	1140
15	CGCTGTTAAG ATAATACCTA ACATGATGAT CATATAACTT AATGAACTCC CTAGGTTTGC	1200
	AACAGGTACT AATGCTGCTC TGAATCTTAA GAACCAATAT CCTTGGTGAT CTTGAATGGC	1260
20	ATGACCAACT TCGTGGGCTG CAATTGCAGT TCCAGCAACT GATGGTCTGT CATAGTTTGC	1320
	AGGAGATAGT GAAACAACCT TCTTTTTAGG ATCGTAATGA TCTGTTAAGA ATCCTTCACC	1380
	TTTAACAAC TCGACATCAT AAATACCGTT TGCATGTAAA ATTTCTAATG CAACTTCACG	1440
25	ACCCGTTTTA CCACTAGTTG ATCTAACTTG TGAATATTTT TCATAGTTAG ATTTAACTTT	1500
	GTGTTGTGCC CATAAAGGAA GCACCATTAA TATTACGAAA TAAATTATCA TAGTAAAAAT	1560
	TGAAGACAAT AAACCTCACTC TCCTTTATAA ATATTTTACT GTCATTTGCC GTTTTATCA	1620
30	AATCATTAC ACTTTAATAA TTTGTTTAA TCAATATAAA GCAAAAGTCC AAAACACTT	1680
	AGACAACATG ATAATACACC AATTTGCCAC ACATGTGTAG TTATAAAATC ATAATATGGA	1740
	AATTGAAGGT GAAAATAGTC AATATAATCA TTCAAAAACA CCCAAATCAT yGCTACACTG	1800
35	ATTCCAATCA TAGAACGTTT AAACCTAGGA TAGAAGTAAA TTGCCTGAAC AGCCATTATA	1860
	CTGTGGGAAA ACATTAATAC CAAACCATT ACTGTAATAT CACCTTGTTT AATAATAAAT	1920
	AATATATTCA TTATACTGC CCAAATCCCA TATTTGAATA ATGTTACAAA TGCCAGTGCA	1980
40	TCGATAATAC TATTTTGT TTGAATTAAAT ATCAATGAGA TAGAAATAAC TAAGTATAAT	2040
	ATTGCAGTTG GGCTATCTGG AACAAAAATC TTAATATGCC AGGGCGTATG ACTTAATTGT	2100
45	TCACCATACC ATATATAACC ATAAATCATC CCTAATATAT TACAAATGAG TAGCATCATT	2160
	AACCAAGAAC GTTGATAAAG TGTATATTGC CAAATGCTT TAATTGTCAT CTGCTAAGTC	2220
	CTCAAATTGA TTATGTTTAT TTAGTAGCTT GAGTGTATTT AAAATTTGCG TTAGTTGATA	2280
50	AAAACGTTGC TTTTCATTCA TCTGTAACT TAAATCAATA TTGTGTAACA AGTAATCTAT	2340
	TAATAACGCA TGTTTATGCC GATCTATAGC CATACTATTT AAGTCATGAA GATAAGTTTG	2400

55

	TGACACGTTT GCGAAGTGAA TTTGAATATC AAAAGCACAG TTATGATTAG CGATATAATC	2520
	AAATATTTCa TTTGTATTCA TTAACTTTAT ATTACGCTTA GTAAATTGAA TTGCAGAAGC	2580
5	GTGACTTCCC ACTTCTGCAA TTTCTAATGT TTCATGATGA TTAATTTTTG TATCTACAAA	2640
	ATGAATGTTT GCCAATTTCTG CCTCATTAC TTTTATATAG TTAAGCACCC AAAGTCAAT	2700
	ACGCGACTTA AATCGATATT GAAAAAGTAA ATATTCAATA AAAGTTTCTT TAATTTGATT	2760
10	GAGTGTCTCT GACATCAAAT ACCCCATTTT AAGATTGCAA TCTTGaTAAT TCGTCATGCC	2820
	AATTTTCGTT ACTTGGcTCT AGTTCCAACA ATTGATTIAA AATAGTAATT GCTTGTTCCT	2880
15	TTTGACCAAT TTCAATTAA TAGAAATAAT AATCACTCAT AAAATCAATA TTTGTTTTCA	2940
	TCGTTGGATA TGCTAATTCA AAGAAATGTT GAGCTTCTTT ATCTCGCTC	2989

(2) INFORMATION FOR SEQ ID NO: 151:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1143 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

	CATCAACTCC TTAATTACAC TGTAATGAT ATGCGTCTTT TTGACAACTA TATTTGTCAA	60
30	ATCTACACCA AAAAAATATGA TTATCCACCT ATGTATGACA TTTTGAAACA AACACCTCAA	120
	CGCCTACAAG TCATAATTGT TTAAGTTTCGT TACACCTTCC TGCATAATTA ACAGCATTCT	180
	AATTTTAGTA TGATGCACGC ATTTTCACTA AATCAAACCA TTCAAAGGAG ACTATTATGG	240
35	CATTTACATT ATCTGCAATT CAACAAGCAC ATCAACAATT TACTGGTGTT GACTTTCCAA	300
	AAGTATTCAA AGCTTTTAAA GATATGGGGA TGACTTACAA TATCGTCAAC ATTCAAGATG	360
40	GCACTGCAAC ATACGTACAT CAATCAGAAG ATGATATCGT TACGTCATCT GTAAAAAGTA	420
	ATCATCCTGT TGCTCAAAAA TCAACAAAA CAATAGTTCA AGACGTCTTA ACTAGACATC	480
	AACAAGGGCA AACAGATTTT GAAACATTTT GTGATGAAAT GGCTGAAGCT GGCATTTATA	540
45	AATGGCATAT CGATATTCmA GCGGGCACTT GTACTTATAT CGACTTGCAA GACCAAGCTG	600
	TTATTTTACA ATTAATCCCT CAATAAATA TATTTATAGC AACATTTTAA TTATTTTATA	660
	AAATTTTATT GATAATCATT ATCGTTCGGT ATAAAGTAAA TACTATATAC TACTTATGAG	720
50	TGAGGTTGAT TATCATGATA ACTAACACTT TTATTTTAGG CATCACAGGC CCAACAAGTC	780
	TTGTCGTCAT TAGCATTATC GCTTTAATTA TTTTGGTCC GAAAAATTA CCACAATTTG	840

AGTCTCACGA TACACCCAGT AAGGAATCGA AACACAGCG AGAGCAATAG CACTGACCAC 960
 ACCTTACTGG TTCACTTTAG CGAACTACGC CATCGGTTAG TAAAAATTTT ATTGTCGTTC 1020
 5 GTCATTACGG TCATCGTCGT ATATGTYTCA TCATTTTGGT GGATGACACC ATTCATAACG 1080
 TATATyACCC GgCACATGTG TcCTTACATG CATTTcATTC ACAGAAATGA TACAAATAAC 1140
 GTG 1143

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7953 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

CAACGCCTGA ACGTAAACCA TATCGTTTCG CGATTTCCCTC ATCTTGACTA TTTACTAAAA 60
 ACTCTCTCAT GCGGATTAAT GTTCTTTTTT CTCTTTTAGT TAATGGTAAT TCTAACTCAG 120
 25 CTGCTTTTTG ACGCAAAGTT GGATGACCAT CTCTAATGAT GTCTTTCATT GTTAACATAT 180
 ATTGCACCTT CCTTATTTTA ATTTGTTTTA GTTGAATGAC AGTAAAAAGG TTGTTAAGAT 240
 ACTCATACAT TTTTATGTGT AAATATCTAC AAAGTTAACC AACTACTGCC AATGTTTATT 300
 30 TTAGATAGTA TATGTAAATT TTCAaGATAT GCgTAATTGC gTTAAAAAAT GaTTAAAGTG 360
 TTGGTTTCAA GCAATGaTAC TTTAGAAATT TATTTATCAT CTTGACTTTA AAAATTATAT 420
 TATAAATGAC GTAACGTGCA ACAGATATAC TTAGTArTGA AGATGTGTAA TGTAATTGTT 480
 35 TAAAATTGAT TTCCAAGCAG ATTTTATTTA TCATTTAATT TAAATAGCAA GTGGAGGTAC 540
 AAGTAATGAA ATTTGGAAAA ACAATCGCAG TAGTATTAGC ATCTAGTGTC TTGCTTGCG 600
 GATGTACTAC GGATAAAAAA GAAATTAAGG CATATTTAAA GCAAGTGGAT AAAATTAAAG 660
 40 ATGATGAAGA ACCAATTAAA ACTGTTGGTA AGAAAATTGC TGAATTAGAT GAGAAAAAGA 720
 AAAAATTAAC TGAAGATGTC AATAGTAAAG ATACAGCAGT TCGCGGTAAA GCAGTAAAGG 780
 ATTTAATTAA AAATGCCGAT GATCGTCTAA AGGAATTTGA AAAAGAAGAA GACGCAATTA 840
 AGAAGTCTGA ACAAGACTTT AAGAAAGCAA AAAGTCACGT TGATAACATT GATAATGATG 900
 TTAAACGTAA AGAAGTAAAA CAATTAGATG ATGTATTAAA AGAAAAATAT AAGTTACACA 960
 50 GTGATTACGC GAAAGCATaT AAAAAGGCTG TAAACTCAGA GAAAACATTA TTAAATATT 1020
 TAAATCAAAA TGACGCGACA CAACAAGGTG TTAACGAAAA ATCAwAAGCA ATAGAACAGA 1080

	AAGAAAAGCA AGACGTTGAT CAATTTAAAT AATTAATATA ATACAGATGG TAGGAAACAA	1200
	CTAATACAGT TCCTATTATC TGTATCTTTT TTTATTAAAA CAGAACTTTT TCAAATGGTT	1260
5	TAACAGTCCC ATTTATTTGT GGTACAATTA GTAAGGATAA AATGAATTTC TATACAATTA	1320
	TGGGAAAGGT ATTGTGAATT GAATGGCTCC TAAGTTACAA GCCCAATTCG ATGCAGTAAA	1380
	AGTTTTAAAT GATACTCAAT CGAAATTTGA AATGGTTCAA ATTTTGGATG AGAATGGTAA	1440
10	CGTCGTAAAT GAAGACTTAG TACCTGATCT TACGGATGAA CAATTAGTGG AATTAATGGA	1500
	AAGAATGGTA TGGACTCGTA TCCTTGATCA ACGTTCTATC TCATTAAACA GACAAGGACG	1560
15	TTTAGGTTTC TATGCACCAA CTGCTGGTCA AGAAGCATCA CAATTAGCGT CACAATACGC	1620
	TTTAGAAAAA GAAGATTACA TTTTACCGGG ATACAGAGAT GTTCCTCAA TTTATTGGCA	1680
	TGGTTTACCA TTAAGTGAAG CTTTCTTATT CTCAAGAGGT CACTTCAAAG GAAATCAATT	1740
20	CCCTGAAGGC GTTAATGCAT TAAGCCCACA AATTATTATC GGTGCACAAT ACATTCAAGC	1800
	TGCTGGTGTT GCATTTGCAC TTAAAAACG TGGTAAAAAT GCAGTTGCAA TCACTTACAC	1860
	TGGTGACGGT GGTTCTTCAC AAGGTGATTT CTACGAaGGT ATTAACTTTG CAGCAGCTTA	1920
25	TAAAGCACCT GCAATTTTCG TTATTCAAAA CAATAACTAT GCAATTTCOA CACCAAGAAG	1980
	CAAGCAAACCT GCTGCTGAAA CATTAGCTCA AAAAGCAATT GCTGTAGGTA TTCCTGGTAT	2040
	CCAAGTTGAT GGTATGGATG CGTTAgcTGT nATATCAAGC AACTAAAGAA GCACGTGACC	2100
30	GCGCagTTGC AGGTGAAGGT CCAACATTAA TTGAACTAT GACATATCGT TATGGTCCCTC	2160
	ATACAATGGC TGGTGACGAT CCAACTCGTT ACAGAAC TTC AGACGAAGAT GCTGAATGGG	2220
35	AGAAAAAAGA CCCATTAGTA CGTTTCCGTA AATTCCTTGA AAACAAAGGT TTATGGAATG	2280
	AAGACAAAGA AAATGAAGTT ATTGAACGTG CAAAAGCTGA TATTAAAGCA GCAATTAAAG	2340
	AGGCTGATAA CACTGAAAAA CAAACTGTTA CTTCTCTAAT GGAAATTATG TATGAAGATA	2400
40	TGCCTCAAAA CTTAGCAGAA CAATATGAAA TTTACAAAGA GAAGGAGTCG AAGTAAGCCA	2460
	TGGCACAAAT GACAATGGTT CAAGCGATTA ATGATGCGCT TAAACTGAA CTTAAAAATG	2520
	ACCAAGATGT TTTAATTTTT GGTGAAGACG TTGGTGTTAA CGGCGGTGTT TTCCGTGTTA	2580
45	CTGAAGGACT ACAAAAAGAA TTTGGTGAAG ATAGAGTATT CGATACACCT TTAGCTGAAT	2640
	CAGGTATTGG TGGTTTAGCG ATGGGTCTTG CAGTTGAAGG ATTCCGTCCG GTTATGGAAG	2700
	TACAATTCTT AGGTTTCGTA TTCGAAGTAT TTGATGCGAT TGCTGGACAA ATTGCACGTA	2760
50	CTCGTTTCCG TTCAGGCGGT ACTAAACTG CACCTGTAAC AATTCGTAGC CCATTGGTG	2820
	GTGGCGTACA CACACCAGAA TTACACGCAG ATAACCTAGA AGGTATTTTA GCTCAATCTC	2880

55

	CTATTAGAAG TAATGACCCA GTCGTATACT TAGAGCATAT GAAATTGTAT CGTTCATTCC	3000
	GTGAAGAAGT ACCTGAAGAA GAATATACAA TTGACATTGG TAAGGCTAAT GTGAAAAAAG	3060
5	AAGGTAATGA CATTTCAATC ATCACATACG GTGCAATGGT TCAAGAATCA ATGAAAGCTG	3120
	CAGAAGAACT TGAAAAAGAT GGTATTCTG TTGAAGTAAT TGAATTACGT ACTGTTCAAC	3180
	CAATCGATGT TGACACAATT GTAGCTTCAG TTGAAAAAAG TGGTCGTGCA GTTGTAGTTC	3240
10	AAGAAGCACA ACGTCAAGCT GGTGTTGGTG CAGCAGTTGT AGCTGAATTA AGTGAACGTG	3300
	CAATCCTTTC ATTAGAAGCA CCTATTGGAA GAGTTGCAGC AGCAGATACA ATTTATCCAT	3360
	TCACTCAAGC TGAAAATGTT TGGTTACCAA ACAAAAATGA CATCATCGAA AAAGCAAAAG	3420
15	AAACTTTAGA ATTTTAATAC ATTTTAAAAG TTAACGAAGT TAGCGTATTT TAGTCTCATT	3480
	GATTAAAATG AAATGTTTAA TTTACGAAAT CTTAGGAGGG CAAAAACGTG GCATTTGAAT	3540
20	TTAGATTACC CGATATCGGG GAAGGTATCC ACGAAGGTGA AATTGTAAAA TGGTTTGTTA	3600
	AAGCTGGAGA TACTATTGAA GAAGACGATG TTTTAGCTGA GGTACAAAAC GATAAATCAG	3660
	TAGTAGAAAT CCCATCACCA GCATCTGGTA CTGTAGAAGA AGTTATGGTA GAAGAAGGTA	3720
25	CAGTAGCTGT AGTTGGTGAC GTTATTGTTA AAATCGATGC ACCTGATGCA GAAGATATGC	3780
	AATTTAAAGG TCATGATGAT GATTCATCAT CTAAAGAAGA ACCTGCGAAA GAGGAAGCGC	3840
	CAGcAGaGCA AGCACCTGTA GCTACTCAAA CTGAAGAAGT AGATGAAAAC AGAACTGTTA	3900
30	AAGCAATGCC TTCAGTACGT AAATACGCAC GTGAAAAAGG TGTTAACATT AAAGCAGTTT	3960
	CTGGATCTGG TAAAAATGGT CGTATTACAA AAGAAGATGT AGATGCATAC TTAAATGGTG	4020
	GTGCACCAAC AGCTTCAAAT GAATCAGCTG CTTCAGCTAC AAGTGAAGAA GTTGCTGAAA	4080
35	CTCCTGCAGC ACCTGCAGCA GTAACATTAG AAGGCGACTT CCCAGAAACA ACTGAAAAAA	4140
	TCCCTGCTAT GCGTAGAGCA ATTGCGAAAG CAATGGTTAA CTCTAAGCAT ACTGCACCTC	4200
	ATGTAAACATT AATGGATGAA ATTGATGTTT AAGCATTATG GGATCACCGT AAGAAATTTA	4260
40	AAGAAATCGC AGCTGAACAA GGTACTAAGT TAACATTCTT ACCTTATGTT GTTAAAGCAC	4320
	TTGTTTCTGC ATTGAAAAAA TACCCAGCAC TTAACACTTC ATTCAATGAA GAAGCTGGTG	4380
45	AAATCGTTCA TAAACATTAC TGGAATATCG GTATTGCAGC AGACACTGAT AGAGGATTAT	4440
	TAGTACCTGT TGTTAAACAT GCTGATCGTA AGTCTATTTT CCAAATTTCA GATGAAATTA	4500
	ATGAATTAGC TGTTAAAGCA CGTGATGGTA AATTAACAGC CGATGAAATG AAAGGTGCTA	4560
50	CATGCACAAT CAGTAATATC GGTTCACTG GTGGACAATG GTTCACTCCA GTTATCAATC	4620
	ACCCAGAAGT AGCAATCTTA GGAATTGGCC GTATTGCTCA AAAACCTATC GTTAAAGATG	4680

55

	ATGGTGCAAC	TGGCCAAAAT	GCAATGAATC	ACATTAAACG	TTTATTAAAT	AATCCAGAAT	4800
	TATTATTAAT	GGAGGGGTAA	AACATGGTAG	TTGGAGATTT	CCCAATTGAA	ACAGATACTA	4860
5	TAGTAATCGG	AGCAGGTCCT	GGTGGATACG	TTGCAGCAAT	TCGTGCAGCT	CAATTAGGAC	4920
	AAAAAGTAAC	AATCGTTGAG	AAAGGTAATC	TTGGTGGTGT	TTGCTTAAAC	GTAGGATGTA	4980
10	TTCTTCAAA	AGCATTACTA	CATGCTTCTC	ACCGTTTTGT	TGAAGCACAA	CATTCTGAAA	5040
	ACTTAGGTGT	TATTGCTGAA	AGTGTTCCTT	TAACTTCCA	AAAAGTTCAA	GAATTCAAAT	5100
	CATCAGTTGT	TAATAAATTA	ACTGGTGGTG	TTGAAAGCTT	ACTTAAAGGT	AACAAAGTTA	5160
15	ACATCGTTAA	AGGTGAAGCA	TATTTCTAG	ATAACAATAG	CTTACGTGTT	ATGGACGAAA	5220
	AGAGCGCACA	AACATACAAC	TTTAAAAATG	CAATCATTGC	AACAGGTTCA	AGACCAATTG	5280
	AAATTCCTAA	TTTCAAATTC	GGTAAACGTG	TTATCGACTC	AACAGGTGCT	TTAAACTTAC	5340
20	AAGAAGTACC	aGGTAAATTA	GTTGTAGTTG	GTGGAGGATA	CATTGGATCA	GAATTAGGTA	5400
	CAGCATTTGC	TAACTTTGGT	TCAGAAGTAA	CCATCCTTGA	AGGTGCTAAA	GATATCTTAG	5460
	GTGGCTTCGA	AAAACAAATG	ACACAACCTG	TTAAAAAAGG	TATGAAAGAA	AAAGGTGTTG	5520
25	AAATCGTTAC	TGAAGCTATG	GCTAAATCAG	CTGAAGAAAC	AGATAACGGA	GTTAAAGTTA	5580
	CTTATGAAGC	TAAAGGCGAA	GAGAAAACAA	TCGAAGCTGA	TTATGTATTA	GTAAGTGTAG	5640
	GTCGTCGTCC	AAACACAGAC	GAATTAGGCC	TAGAAGAATT	AGGTGTTAAA	TCGCTGACC	5700
30	GTGGATTATT	AGAAGTTGAT	AAACAAAGCC	GTACGTCTAT	CAGCAATATC	TATGCAATTG	5760
	GTGATATCGT	TCCAGGTTTA	CCACTTGCTC	ACAAAGCTAG	CTATGAAGCT	AAAGTTGCTG	5820
	CTGAAGCAAT	TGATGGTCAA	GCTGCTGAAG	TTGATTACAT	TGGTATGCCA	GCAGTATGCT	5880
35	TTACTGAACC	AGAATTAGCT	ACAGTTGGTT	ATTCAGAAGC	GCAAGCTAAA	GAAGAAGGTT	5940
	TAGCAATTAA	AGCTTCTAAA	TTCCCATATG	CAGCAAATGG	TCGTGCATTA	TCATTAGATG	6000
40	ATACTAACGG	ATTTGTTAAA	CTTATTACAC	TTAAAGAAGA	TGATACTTTA	ATCGGTGCTC	6060
	AAGTAGTTGG	TACTGGTGCA	TCAGATATTA	TCTCTGAATT	AGGTTTAGCA	ATTGAAGCTG	6120
	GTATGAATGC	TGAAGATATC	GCATTAAACA	TCCATGCACA	TCCAACATTA	GGTGAGATGA	6180
45	CTATGGAAGC	AGCAGAAAAA	GCTATCGGAT	ACCCAATCCA	TACAATGTAA	TAACTGATTA	6240
	TCTATAAAGA	TTCAGTCATT	AAAAGCTGTA	GCATATGCTA	CGGCTTTTTT	GTTTTAGGTA	6300
	AAGTAATGTA	AGGAAATTGA	TTTGAGATAT	CGTTAACATG	TGACATGCAT	GTTATACTAG	6360
50	CGATGCTAAT	AAAAGAATTG	AAATGGAGGG	TTCAACAATG	GAATATGAGT	ATCCAATTGA	6420
	TTTAGACTGG	AGTAATGAAG	AGATGATTTT	AGTGATAAAT	TTCTTTAATC	ATGTAGAGAA	6480
55							

AATTGTGCCT GCTAAAGCAG AGGAAAAACA AATTTTAAAT ACTTTCGAAA AAAGTAGTGG 6600
 CTATAATAGT TACAAAGCAG TTCAAGATGT AAAAAGTCAC TCTGAAGAAC AAAGAGTAAC 6660
 5 AGCTAAAnAA TAATTCGTTT GAAATTAACA CAATTTAATA GGAATTTTTC TTTAAACTA 6720
 TTGCTAATAA AGCTATATTT TGATACCTTT ATCAAGTGTT AAACAAAATG TTTGATAAAA 6780
 GTAAACTTAA TATAGCTTTT TTAGGTGGAA AAATAAATGA ACATAGGTAA TAAAATTAAA 6840
 10 AATCTTAGAA GAATTAAAAA TTTAACGCAA GAAGAACTTG CTGAACGTAC AGACTTATCG 6900
 AAAGGCTACA TTTCACAAAT AGAAAGTGAA CATGCCTCAC CAAGTATGGA AACTTTCTTA 6960
 AATATTATAG AGGTGTTAGG AACGACGCCA AGTGAATTTT TTAAAGACAG TGAAAATGAA 7020
 15 AAAGTATTAT ACAAGAAGGA AGAACAAGTT ATTTATGATG AGTATGATGA AGGTTATATA 7080
 TTAAATTGGT TAGTTTCAAA GTCAAATGAA TATGATATGG AGCCATTAAT ATTAACTTTA 7140
 AAGCCTGGAG CATCATATAA AAATTTTAAAT CCATCAGAGT CTGATACGTT TATTTATTGT 7200
 ATGTCAGGTC AGATAACACT TAATTTAGGC AAAGAGATAT ATCAAGCACA AGAAGAAGAC 7260
 GTTTGTATT TTAAAGCAGC AGATAATCAT CGTTTGTCAA ACGAATCAA CAATGAAACA 7320
 20 CGAATACTTA TTGTAGCGAC AGCTTCATAT TTATAGGGGG GATCTTATTT GGAACCGTTA 7380
 TTATCATTAA AATCAGTTAG TAAAAGCTAT GATGATCTTA ATATCTTAGA TGACATAGAT 7440
 ATTGATATTG AATCAGGATA CTTTATACA TTATTAGGTC CTTCAGGTTG TGGTAAAACA 7500
 30 ACAATTTTAA AATTAATTGC AGGGTTTGAA TATCCTGACA GTGGTGAAGT GATTTATCAA 7560
 AACAAACCAA TTGGTAATTT ACCACCAAAT AAACGTAAAG TGAATACAGT CTTTCAAGAT 7620
 TATGCATTAT TTCCACACTT AAACGTCTAT GATAATATCG CTTTGGTTT GAAATTAAAA 7680
 35 AAATTATCAA AAACCGAAAT TGATCAAAAA GTAAGTGAGG CATTAAAATT AGTAAACTT 7740
 TCAGTTATG AAAAAAGAAA TATTAATGAA ATGAGTGGCG GACAAAAGCA ACGTGTGCA 7800
 ATTGCACGTG CTATCGTAAA TGAACCAGAA ATATTATTGT TAGATGAATC TTTATCCGCA 7860
 40 TTAGATTGTA AATTGCGTAC TGAAATGCAA TATGAATTAC GAGAATTGCa ATCTAGATTA 7920
 GGTATTACAT TTATATTTGT aACACATGAT CCA 7953

45 (2) INFORMATION FOR SEQ ID NO: 153:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2347 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

55

	GGCGTGATCA TACGACCGTC ATTCATGCTC ATGAAAAAAT ATCTAAAGAT TTAAAAGAAG	60
	ATCCTATTTT TAAACAAGAA GTAGAGAATC TTGAAAAAGA AATAAGAAAT GTATAAGTAG	120
5	GAAACTTTGG GAAATGTAAT CTGTTATATA ACAGCACTAA TGATnACAAT CATTITTTTAC	180
	ATTTCTATAT GCTAATGTGG CAAGATGAGC AAAACTCATT TTGTGGATaA TGTTTAAAAG	240
	TCATACACAC CATAACAAG TTATCAACAT GTGTATAAyT cGcCAAATCT ATGTTTTTAA	300
10	GACTTATCCA CCAATCCACA GCACCTACTA CTATTACTAA GAACTTAAAA CCTATATAAT	360
	TATATATAAA CGACTGGAAG GAGTTTTAAT TAATGATGGA ATTcACTATT AAAAGAGATT	420
	ATTTTATTAC ACAATTaAAT GACACATTAA AAGCTATTTT ACCAAGaACA ACATTACCTA	480
15	TATTAAGTGG TATCAAAATC GATGCGAAAG AACATGAAGT TATATTaACT GGTTCAGACT	540
	CTGAAATTTT AATAGAAATC ACTATTCCTA AAAGTGTAGA TGGCGAAGAT ATTGTCAATA	600
20	TTTCAGAAAC AGGCTCAGTA GTACTTCCTG GACGATTCTT TGTTGATATT ATAAAAAAT	660
	TACCTGGTAA AGATGTTAAA TTATCTACAA ATGAACAATT CCAGACATTA ATTACATCAG	720
	GTCATTCTGA ATTTAATTTA AGTGGCTTAG ATCCAGATCA ATATCCTTTA TTACCTCAAG	780
25	TTTCTAGAGA TGACGCAATT CAATTGTCGG TAAAAGTGCT TAAAAACGTG ATTGCACAAA	840
	CAAATTTTGC AGTGTCCAcC TCAGAAACAC GCCCAGTACT AACTGGTGTG AACTGGCTTA	900
	TACAAGAAAA TGAATTAATA TGCACAGCGA CTGACTCACA CCGCTTGGCT GTAAGAAAGT	960
30	TGCAGTTAGA AGATGTTTCT GAAAACAAAA ATGTCATCAT TCCAGGTAAG GCTTTAGCTG	1020
	AATTAAATAA AATTATGTCT GACAATGAAG AAGACATTGA TATCTTCTTT GCTTCAAACC	1080
	AAGTTTTATT TAAAGTTGGA AATGTGAACT TTATTTCTCG ATTATTAGAA GGACATTATC	1140
35	CTGATACAAC ACGTTTATTC CCTGAAAACCT ATGAAATTA AATTAAGTATA GACAATGGGG	1200
	AGTTTATCA TGCGATTGAT CGTGCCTCTT TATTAGCGCG TGAAGGTGGT AATAACGTTA	1260
40	TTAAATTAAG TACAGGTGAT GACGTTGTTG AATTGTCTTC TACATCACCA GAAATTGGTA	1320
	CTGTAAAAGA AGAAGTTGAT GCAAACGATG TTGAAGGTGG TAGCCTGAAA ATTTCAATCA	1380
	ACTCTAAATA TATGATGGAT GCTTTAAAAG CAATCGATAA TGATGAGGTT GAAGTTGAAT	1440
45	TCTTCGGTAC AATGAAACCA TTTATTCTAA AACCAAAAGG TGACGACTCG GTAACGCAAT	1500
	TAATTTTACC AATCAGAAGT TACTAAAAAT AAATATAAAT AAAGGATGAC GTGATTAATT	1560
	AAAACGTCAT CCTTTATTTT TTGGCAAAAA TAATTCTAGG TGCGTATGTA AAATAAATTT	1620
50	GGCAGCATTT TAAACAGCAA ATAAAAGACG CCAATTAAAT TTATGACAAA TGTATCCAAA	1680
	ATTTAATAAG TGTGCTTATA TGCCCTTTAA ATTTAAAATT TTAATAGTCA ATAACAAGTT	1740

55

AAAAATAAGA ATTAATTATT TATATGTAAA CGGTTTCTAC CTCTATTTTA AATGAAATTT 1860
 GTGACAAAAA AAGGTATAAT ATATTAATGA CATACAAAGA AATGGAGTGA TTATTTTGGT 1920
 5 TCAAGAAGTT GTAGTAGAAG GAGACATTAA TTTAGGTCAA TTTCTAAAAA CAGAAGGGAT 1980
 TATTGAATCT GGTGGTCAAG CAAAATGGTT CTTGCAAGAC GTTGAAGTAT TAATTAATGG 2040
 AGTGCCTGAA ACACGTCGCG GTAAAAAGTT AGAACATCAA GATCGTATAG ATATCCCAGA 2100
 10 ATTACCTGAA GATGCTGGTT CTTTCTTAAT CATTTCATCAA GGTGAACAAT GAAGTTAAAT 2160
 ACACTCCAAT TAGAAAATTA TCGTAACTAT GATGAGGTTA CGTTGAAATG TCATCCTGAC 2220
 GTGAATATCC TCATTGGAGA AAATGCACAA GGGAAAGACA AATTTACTTG GAATCAATTT 2280
 15 ATACCTTAGC TTTAGCAAAA AGTCATAGAA CGAGTAATGG ATAAGGGACT CCATACCGTT 2340
 TTAATGC 2347

20 (2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13542 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:
 30 ACAAGACGTh TCTATAACTT ATCTGAAATC GCTCGTCAAG ATAAAGATTA TGCAACTATC 60
 TCATTCTTAA ACTGGTCTT AGATGAACAA GTCGAAGAAG AATCAATGTT TGAAACTCAC 120
 ATCAATTATT TAACTCGTAT CGGCGATGAC AGCAATGCAT TATATCTTTA CGAAAAAGAA 180
 35 CTTGGCGCTC GTACATTCGA CGAAGAATAA TTAAACATCA CTACAATAGA CAGATAAATA 240
 TCATACGACA TGATAGGCAT TTGGGTCACT TACAATAACC CAATGTCTAT ATTATTTTGC 300
 TTTACGGAGA TCACTAGATT CATTTTCTGA ATCATTGATC TCGTTTTTTT CATTTTCAAG 360
 40 GCTAATTATT GTATTTTTAG TCATTTATTT TTTAAACTAC TAATGTTAAT AACTCTAAAT 420
 TTGATGTTGA ATTAATTTGA CGATTTTAAA GCATATCATC ATTTACTTTT TAATCAGAGT 480
 TACATCCAAA TGATAGATTT CACGTTATAC CTTACGTAT AATATTATGT ATCGTTTGTA 540
 45 AGCAAATGAC TAAAAGTCTA TTAATATATA CATTTAATTA ATTGAAAGGA TTGACTACAT 600
 GATACAAGAT GCGTTTGTG CACTTGATTT TGAAACAGCA AATGGTAAAC GTACAAGTAT 660
 50 TTGTTCTGTG GGAATGGTTA AAGTCATTGA TAGTCAAATA ACAGAAACAT TTCATACTCT 720
 TGTGAATCCG CAAGACTATT TTTACAACA AAATATTAATA ATTCATGGCA TACAACCAGA 780

55

	aGATTTACCT GTTGTGCGAC ATAACGCGGC ATTTGATATG AACGTCTTAC ATCAAAGCAT	900
	TCAAATATT GGTTTACCAA CTCCAAATTT AACTTACTTT TGTAAGTTATC AACTTGCTAA	960
5	AAGAACCGTT GATTCGTATC GATACGGTTT AAAACATATG ATGGAGTTTT ATCAATTAGA	1020
	TTTTCATGGT CATCATGATG CATTGAATGA TGCCAAAGCA TCGCAATGA TTACTTTTAG	1080
	GCTACTGAAA AATTATGAAA ATTTAACATA TGTAACATAAT ATTTATGGTA AAAATCTAAA	1140
10	AGATAAAGGC TAGGACTAAA TAAAATACTC CCTTCAAAAG TAAGCATTGT AAAAATGTAA	1200
	ACTTTGCAGG GAGCTTTATT TTATATAAAG TCATATATCG TCATATTTTT ATAAGTTGAT	1260
	TGTTCTAAAT TACCTACAGT GACACCAATA AGTCGAATTG GTACATCAGG GTCTTTTAAA	1320
15	TCGTTATAAA GTAAATATGC AATATTATAA ATATCTTCTT CAGAACTAAC CGAATCTCTT	1380
	AAACTCATCT GTTTAGATAG CGTTTCAAAT TGATAAGTTT TAATTTTAAC CGTTACAGTT	1440
20	TTAGCTGACT TCTGTAATTT ATTTAGACGT TCAGCTGTTT TACCTGnACA ATTCCCATAC	1500
	TTTTCTTAAA ATCTCTTCAT CATCATTCAC GTCTGTTGCA AATGTGCGTT CAGTCCCTAC	1560
	TGATTTCTT ACTCTTGATG ATTTCACTTC ACTATGGTCA ATACCGCGTG CCTTGTATA	1620
25	TAAACCCCGA CCTCTTTTTC CAAACAAACG TATTAATTCA AATTCCGTTT TCTCATATAA	1680
	ATCTCTACCG TTAATAATAC CATTATCATG CATTACTTTT TTGGAAGCTT TACCTACGCC	1740
	TGGaAAATCT CCAATATCCA ATGTCATCAA AATATCATGG aCATTTTGAT AATCAATCAC	1800
30	AGTCATACCA TCAGGTTTAT TCATACCACT CGCTAATTTA GCTAAAAATT TGTTATAAGA	1860
	AACACCTGCA GATGCTGTTA AATGTGTCTG CTCTAGAATA TCTTTTCTAA TATACTGAGC	1920
	AATTTTCGAA GCAGGAAGGT CTGGTCTCAC TAATTCTGTA ATATCTAAAT ACGCTTCATC	1980
35	CAATGACATC GGTTCACCT TATCTGTATA ACTTCGGAAA ATAGACATAA TCTGCGCAGA	2040
	TGTTTCTCGG TAAGCACCAA AATTACTTGT GACAAAGTAT CCATTTGGAC ATAATTATG	2100
	CGCTGTGAC ATAGGCATTG CTGAATGGAC GCCGTATTTT CGTGCTTCAT AGGATGCCGT	2160
40	AGAGACAACA CCCCTACTGC TTGCTTTACC ACCAACAATG ACTGGTTTCC CTTTCAATTT	2220
	GGGGTTATCT CTCATTTCGA CTTGTGCAAA AAAATAGTCC ATATCTATAT GAATAATTCG	2280
45	TCTCTCAGTC AAGTGCTCAC CTCCCTACTA ATTTTACTT TTATAACGCA CAAAAATATC	2340
	TCAACATAAT TATACGCTGT GTACGATTTT TTTACATAAA TCTTGCACTT AGCGATAACT	2400
	ATATTGaGAT AACTACAAGT TGTTATaAAA TCAATTGCTA TTTAAGCATG ATGATGAAGA	2460
50	CGATTGAGTA AGAAAACATA GGTAATCTGA AATAATTCAA GCAAATTCAT TTTGTTGGTA	2520
	TCATCATATT AAAATTTATT ATTGAGTCGG CTTTGTATGA TACAAATAAA TACTATCTTC	2580

55

	AAAGCAATAA	GCGGTATGCA	TACTAAACAT	AAAAATAAGT	GATGAATAAC	CAAATACCTT	2700
	AATTAAAATA	AGCAAGCCAG	TACTTAATAG	GATTAGTGGT	GACAGCATAA	TAATTGAGAA	2760
5	TTGCCATTTG	TTGAAGCAAG	CATCTGCTGT	TTGGAATAAG	ATTCTGTCTT	TTTTTATATT	2820
	AAACATAGGT	TTGCTATCTT	TTTTAAATAA	AAGAAATAAT	GCTCTATGGA	TAAGTTCATG	2880
	TAAAATCAAT	AAAATAATGA	ATCCAGCAAA	CCCATATACA	AGATTGATGA	TGATATTTTG	2940
10	ATCGACAACC	GCTGTGACAC	CTAACGCCCA	CTTATACGTA	AATAAAATCA	CGAATAACGC	3000
	AATAACAAGT	TGCAAGATAA	TAAACCTTCG	CATTTGAAAA	TTATTTGTCTG	TTAAATCAAT	3060
	TTTATGCATT	ACCAACCCTC	CCGATCATGA	CATTCTTATT	CTTCTTTAAA	TATAGTATAC	3120
15	AATGTCACAT	TTAATTTTAA	AAGTTCATAT	CAAGAAAGTA	AATTGGCTGT	AATAAAATTT	3180
	TAATATACGA	CTTCTTTCTT	CACTTATTAA	GGCGAAATTT	TATCTCAAAT	CATGTGCGCT	3240
20	ATTTCAAATT	GAATAATGCC	ACTGTCTCAA	CATGTGTTGT	TTGTGGAAAC	ATATCTACCG	3300
	GTGTTACCTC	TTCAAGTTGA	TATTTTTCAG	CTAATAATAA	TGCATCACGT	TGCTGTGTTG	3360
	CGGGATTACA	TGAAATATAG	ACAATACGCT	TAGGTTCTAA	TGTAAGCAAA	GTCTGAATAA	3420
25	ACGTTTCGTC	ACAGCCCTTT	CTTGGCGGAT	CAACCATTTAC	AACATCTGGT	TTAATCCCTT	3480
	GTGCTTTCCA	TTGTAAAATA	ACTTCTTCAG	CTTTCCCACA	GACAAAAGTT	GTATTATTGC	3540
	ATTGGTTTAT	AGTCGCATTT	TGTTGTGCGT	CTTCAATTGC	AGAAGGTACT	ACTTCAACAC	3600
30	CGTATACATG	TTTTCGAAGT	GGTGCCATAT	ATAGCCCTAT	TGTTCCAATA	CCACAATAGG	3660
	TATCTAATAC	AACTTCATTA	CCTGTCAATT	GCGCATACTC	AATTGCTTTA	TTATATAATT	3720
	TCTCTGTTTG	TTCAGAATTA	ATTTGGTAGA	ATGACTGATC	ACTTATTTTA	AATGTACTAT	3780
35	CTGTTAATTG	ATCAATAATT	GTATCTTTAC	CATATAGCGT	TATAGATTGA	CGTCCCATAA	3840
	TAACATTAGA	GTGGCTATCA	TTAATGTTTT	GTTTAATGCT	TGTCACATTA	GGAAATGCAT	3900
	CTAATATCTT	CTCAACAACA	GCATTTTTTT	GTGGCCACTT	TTTACCATTA	GTTACAAAAA	3960
40	TAATCATCAT	TTCGTCTGTA	TGATATCCTG	TTCTTACAAC	CAAATGTCTC	ATTAAACCTT	4020
	TTTTCAATTG	TTCTTGATAA	ATACTTACAT	TTAAATCTTT	TAAAATAGAT	TTAACTTCAT	4080
45	TCATCACTTC	TTGATGTTGT	GAATCTTGTA	TTAAACAAC	TTCCATGTCA	ATAATGTCAT	4140
	GGCTTCTTTG	ACGATAAAAG	CCCATAATAA	CTTCATTCTG	TTCAATCTTA	CCAACTGGAA	4200
	TCTGGGACTT	GTTTCGATAT	CTCCAAGGAT	CTGTCATGCC	AACTGTATCG	TTAATCTTAG	4260
50	AATTATCAAA	ATGCGCTTTT	CGCTGAAACA	AATTAATCAC	TTGTTCCCTT	TTCAATTCAA	4320
	GTTGTGCTTC	GTATGATAAG	TGTTGAAGTT	GGCACCAC	ACAACGTTCA	TAATATATAC	4380

55

	AGTTCCTTTT TACTTTGATA ATTTTATATT CAATTTGTTC ATTAATTAAA GCTTGTGGTA	4500
	TGAAAATAGG AAAGCGATCT ATTTTACGA CACCATGGCC TTCATGCGTT AAATCAACAA	4560
5	CTGTTCCCGT TTTTATGTCA TTTTLAGCTA TTGCTTGCAA AATTTTACCT CCAAAATGAA	4620
	CAGGTTAGGA ACAAAATTAT GCGCTTCCTA ACCTGCCATT ATATATTTCA CTATTTCTGT	4680
	TTATCTTCG ATTAAATTGT CATCAACATG ATCATTATTT ATTAACCTCTT CATTTACAAT	4740
10	ATCATTAGGT GCAAAGACAT CTATATGACG TTCTAGATTT AAGAAATTCG CTGGTAATTT	4800
	ACCACCATAT TCTCCATCTA CATTTAGTTG TAAGTCTGTG AATGATGAAA TATTAATTGC	4860
	CTTTGCTTTT TCATAAATAA CTTTAGGATG CTTAGTATGT TCTCCTCTTG AAGCTAAAGT	4920
15	CATAATATGA CCAAGTTCTG CAAGGTTTGA TTTTCAACT ATAATTAACG TAAATAGCC	4980
	GTCATCTAAC TTAGCGTCCG GCACTAATTT TTCAAATCCT GCCATTGAAT TTGTTAAACC	5040
20	TAAAAAGAAT AATAATGCTT CTCCTTGGAA AACATTACCA TCATATTCAA TTCTTAAATC	5100
	TACAGCTTTC ATTTGAGGTA ACATTTTCGAA ACCTTTGATG TAATAAGCAA ATGGACCAAC	5160
	AATAGATTTT AATTTACTCG GTGTTTCATA AGAGACTTGC GTCAATTGTC CGCCTGCAGC	5220
25	TAAATTAATA AAGTATCGAT TATTCATTTT ACCAATATCT ACTTTAGTAG AATGACCTTC	5280
	AATGATGACA TCAAGTGCCC CCATGATGTC ATTAGGTATA TGCAATGCAC GTCCAAAGTC	5340
	ATTAACAGTA CCCATAGGAA TGACACCTAG CTTAGGACGA TTAGGCTTTT CTGCGATACC	5400
30	ATTAACACTT TCATTTAATG TTCCATCACC ACCTGCAGCG ATTAATACAT CATAATTTTC	5460
	ATGCATAGCT CTTTCTGCTT CAAGTGTGGC ATCACCTATT TTCTCGGTTG CATATGCACT	5520
	CGTTTCATAT CCCGCTTTT CTAATTTTAT TAAGGCATCA GGTAATTCTC TTTTAAATAG	5580
35	CTCTTTACCT GATGTCGGGT TATAAATGAT TCTAGCACGT TTCCTCATAT CTTATCCCTC	5640
	TACTTAAAT TCATATATTT TAACCTCATC TTTGTTTCGT CTAATAGGGA GTGGGACAGA	5700
	AATAATATTT AACAAAATT ATTTGTTCT ACCCCAACCT GCATTGTCTG TAGAATTTCC	5760
40	TTTCGAAATT CTCTATGTTG GGGCCCCACC CCAACTTGCA CATTATTGTA AGCTGACAGA	5820
	AAGTCAGCTT CTTTGTGTTG GGGCCCCGCC AACTTGCACA TTATTGTAAG CTGACAGAAA	5880
45	ATCAGCTTCT ATGTTGGGGC CCCACTAGAA TTGAAAAAAG CTTGTTACAA GCGTATTTTC	5940
	TTTCAGTCAA CTACAGCCAA TATAACATTG TAGTGCCTAG GACATTGAAT TTATGACCCA	6000
	GGCTCAGTCT TATTTTCATCA TTCTTAATAT CGTTAAAGAC CAACTTGTAT CTTAAACAAA	6060
50	TACTATCTCA ATATGTACAA AGCTTGTAT TTATTCAGCA TTTTGTGCCG TTCTTCATTA	6120
	TATAGCTTCG TCAGTTATGC TATTTTACCT TTAAAAATGAT GTTGTAATA TAATGTTGTC	6180

55

	AACGCATTAA TAAAATTAAT ATTTTACCA TTAACATGTA CAATGAATAA AGTTAAAAGT	6300
	AATTTGACTT CTATAGATAT AAATAAACCC TCGATTGCAT CTAAGTCAGC AATCAAGGGT	6360
5	TTATTTTTTA AATCTTCATA GTTTGATGAT TTAAATTATC TTTTATCTAA TTCTTGTTTT	6420
	AATAGTTGAT TTAATAATTG TGGATTAGCT TGACCTTTAG ACGCTTTCAT AATTTGACCA	6480
10	ACTAAGAAGC CCATAGCTTT GCCTTTACCA TTTTGTAAAT CTTCAACTGA TTGTTGTTA	6540
	TTGTCTAATG CTTCAATTTAC AAATTTTAGA AGTGTGCTT CATCAGAAAT TTGAACTAAG	6600
	CCATTATCTT CCATAATCTG TTTAGCATT CCACCTTTAG CTGCTAACTC TGGGAAGACT	6660
15	TTCTTCGCAA TTTTACTGCT CATTGTTCCG TCTTCGATAA GTTTAATCAT ACCTGCTAAA	6720
	TTTTCTGGTG TTAATTTAGT ATCTAATAAT TCTACTTGAT TTTTATTTAA ATATTCGTTT	6780
	ACGCCACCCA TTAACCAGTT AGATGTTAAT TTAACATCTG CACCGTGTTC AATGTTGAT	6840
20	TCAAAGAAAT CTGACATTTT TTTAGTCAAT GTTAATACGT GTGCATCGTA TGCAGGTAAA	6900
	CCTAATTCAT TTACATACTT AGCTTTACGT TCATCTGGTA ATTCAGGAAT TGTCTGACGA	6960
	ACACGCTCTT TCCAAGCATC ATCAATATAT AAAGGTACAA TGTCAGGCTC TGGGAAGTAA	7020
25	CGGTAATCAT CAGAACCTTC TTTAACACGC ATTAATAATTG TTTTACCTGT AGATTCATCA	7080
	AATCGACGTG TTTCTTGTCC GATTTCTCCA CCATTTAACA ATTCTTCTTC TTGGCGTTTT	7140
	TCTTCATATT CTAAACCTTT ACGTACATAG TTAAATGAGT TTAAGTTTTT CAATTCGGCT	7200
30	TTAGTACCAA ATTTTTCTTG ACCATATGGA CGTAAAGAGA TGTTAGCATC ACAACGTAAA	7260
	GATCCCTCTT CCATCTTAAC GTCTGATACA CCAGTGTATT GAATAATTGA ACGCAATTTT	7320
35	TCTAAATATG CATATGCTTC TTTAGGTGAA CGAATATCTG GTTCAGATAC GATTCAATT	7380
	AGCGGTGTAC CTTGACGGTT CAAGTCAACT AATGAATACT CACCTTTATG TGTGACTTA	7440
	CCAGCATCTT CTTCCATGTG AAGACGAGTA ATACCGATTC GTTTTGTTC ACCGTCGACT	7500
40	TCGATATCGA TATATCCATT TTCACCAATT GGTGATCAA ATTGAGAAAT TTGATATGCT	7560
	TTTGGATTAT CTGGATAGAA ATAGTTCTTA CGGTCAAACCT TAGATTCTGT TGCGATTTCC	7620
	ATATTTAGTG CCATTGCAGC ACGCATTGCC CAGTCTACTG CACGCTTATT AACAACTGGT	7680
45	AAGACACCTG GATATGCTAA GTCGATAACA TTTGTATTG AGTTAGGTTT TGCTCCAAAA	7740
	TGCGCTGGTG ATGGAGAAAA CATTTTTGAG TCCGTTTTTA ACTCTACGTG AACTTCAAGT	7800
	CCTATAACTG TTTCAAAATG CATGATTTCC ACTCCTTATA ATTTTTCATA AACGTCATGT	7860
50	AAATTGTATT GTGTTTCATA TTGATAAGCG ACACGATATA ACGTTTTTTC ATCGAATGGT	7920
	TTACCAATGA ACTGTAAACC GATTGGTCGG CCATTTGATT GTCCACAAGG AACAGAAATA	7980

55

	GGATCATCAA TTTCTTCACC TAAATTA AAC GCaGTgTrAG GCGCTGTTGG ACCAACTACT	8100
	ACATCATAAT TTTTGAATAC TTTATCAAAG TCATTTTTTAA TCAATGTTCT AACTTTTTTGA	8160
5	GATTTTTTAT AGTAAGCATC ATAGTAACCT GAACTTAATG CAAATGTACC TAAGAAAATA	8220
	CGACGTTTTA CTTCTTTACC GAAACCTTCA GATCTTGACA TTTTATATAA TTCTTCTAAT	8280
	GAATGAGCTT CTTTAGAATG ATAACCATAA CGAATTCCGT CAAAACGAGA AAGGTTTGAC	8340
10	GAAGCTTCTG ATGATGCAAT CACGTAATAT GATGGAATAC CAAATTTAGT ATTTGGCAAT	8400
	GATACTTCCT CAACGACAGC ACCTAAAGAT TTAAAGTTT CTACAGCGTT TTGAACTGCT	8460
	TCTTTTACGT CATCAGCTAC ACCTTCACCT AAGTATTCTT TAGGTAATGC AACTTTTAAT	8520
15	CCTTTAATAT CTTTACCAAT TTCAGATGTA AAGTCTACAT CATCAACTGG TGCACCTGTA	8580
	GAGTCATTAA CATCTGCACC AGAAATAGCT TCTAATACGA TTGCATTATC TTTTACATTT	8640
20	CGAGTCAATG GACCAATTTG GTCTAATGAA GATGCAAAAG CAACTAATCC AAATCGAGAT	8700
	ACACGACCGT ATGTTGGTTT CATACCGACA ACGCCACAAT ATGCAGCCGG TTGTCTAATT	8760
	GAACCACCTG TGTCTGAACC TAAGCTAAAT GGTACTAAGC CAGCTGCAAC TGCTGCTGCA	8820
25	GATCCACCTG ATGAACCACC TGGCACTGCT TTATGGTCAA ATGGGTTAAC TGTTTTTTTG	8880
	AAATAAGATG TTTCTGTTGA ACCACCCATT GCAAACCTCAT CCATATTTAA TTTACCGATT	8940
	AAAACGGCAT TTTCAATTATG TAGTTTTTCC ATTACAGTAG ATTCGTAAAT TGGCACAAAA	9000
30	CCTTCTAACA TTTTACTTGC ACATGTTGTT TCTAATCCGT TTGTAATAAT GTTATCTTTT	9060
	ATACCCATTG GAATACCAAA TAATTTGCCA TCCATTTGAT CTTTGTCTTG TAATTCATCC	9120
	AATTCCTGCG CTTTTTTGAT TGCATTTTCT TTATCCAGCG CTAGAAAAGA CTTAATTGTT	9180
35	GGATCAGTCT CTTCAATTGC ATCATATATA TCTTTAACAA CATCAGATGG TTTGATTTTT	9240
	TTGTCTTTTA TTAAAGTTAA TAAATTCTCA ACCGATTCGT AGCGAATGCT CATCTTACGC	9300
40	GTCCTCCTCA TTCATGATTG TAGGCACTTT AAATTGTCCA TCTTCTGTTT CTTTGGCATT	9360
	TTTCAAAGCT AATTCTTG TGGAATACCTTT AATTGCTTTA TCTTCACGTA AAACGTTTTG	9420
	TAAATCTAAA ACGTGATATG TAGGTTCAAC GCCTTCTGTA TCAGCGCTAT CATTTTGTTT	9480
45	TGCAAAATCT AAAATGCTTT CTAATGTGTT GGCCATTTCT TCCGTTTCTT CAGGAGAAAT	9540
	TTGAAGTCTT GCAAGATTCTG CGATATGCTC AACTTCTTCA CGTGTTACTT TTGTCATTAA	9600
	TAAAAGCCTC CTTTAAGTCA TTCATCACTA AATTGTATCA AATTCCAAT TAAAAATCTA	9660
50	AGTATTTATG AGGTGCTACT TTAATTTTCA TAAACTGTA TAAACATTAT CATTCGTTTA	9720
	TCAAATCATT TTTTATGAAA ACAACACTCT TTTAATATTA GACAACCCAA TTCAATATTA	9780

55

TATATTGGTA TGCAAGTATT TCAAAAAGAA TAAATTTAAT TTTCCTACTT TTCTAAACAT 9900
 TTATCTTTAT GTATAATGTT TTCAAGTAAC TAAATTATAA ATTAAATAAA GGGAGTGTTT 9960
 5 ATCATGCTTA CAATGGGGAC AGCATTAAAGT CAACAAGTAG ATGCCAATTG GCAAACCTTAT 10020
 ATTATGATTG CCGTCTACTT CTTGATACTA ATCGTTATTG GCTTTTACGG TTACAAGCAA 10080
 GCAACTGGTA ACCTAAGCGA GTACATGTTA GGTGGACGTA TATTGGACCG TATATTACTG 10140
 10 CATTATCAGC TGGAGCTTCA GATATGAGTG GATGGATGAT TATGGGGCTA CCTGGTTCTG 10200
 TCTATAGCAC TGGTCTATCA GCTATGTGGA TTACAATCGG TTTAACATTA GGTGCTTATA 10260
 TAAATTACTT TGTGTGTTGCT CCTAGACTTC GTGTTTATAC CGAATTAGCT GGAGATGCAA 10320
 15 TTACATTACC AGATTTCTTT AAAAATCGTT TAAACGATAA AAATAATGTG TTAAAGATTA 10380
 TTTCTGGATT GATTATCGTA GTATTCTTTA CATTATATAC ACATTCTGGT TTCGTATCTG 10440
 GTGGTAAACT ATTTGAAAGT GCTTTTGGAT TAGATTATCA TTTCGGTTTA ATATTAGTTG 10500
 20 CTTTCATTGT CATTTTCTAT ACTTCTTTG GTGGATATTT AGCTGTATCA ATTACAGATT 10560
 TCTTCCAAGG TGTCAATTATG TTAATTGCGA TGGTTATGGT CCCTATTGTT GCTATGATGA 10620
 ATTTAAACGG CTGGGGAACG TTTCATGATG TAGCAGCTAT GAAACCTACA AATTTAAATT 10680
 25 TATTTAAAGG GTTATCATTT ATAGGAATTA TCTCTCTATT TTCATGGGGA TTAGGTTATT 10740
 TCGGTCAACC TCATATCATT GTAAGGTTTA TGTCTATTAA ATCACACAAG ATGCTACCTA 10800
 30 AAGCTAGACG TTTAGGTATT AGCTGGATGG CTGTTGGTTT ATTAGGCGCT GTGGCTGTTG 10860
 GTTTAACAGG TATTGCATTC GTACCTGCTT ATCATATTAA ACTAGAAGAT CCTGAGACAT 10920
 TATTCATCGT GATGAGTCAA GTACTCTTCC ATCCTCTTGT AGGTGGTTTC TTACTTGCTG 10980
 35 CGATTCTAGC TGCAATTATG AGCACGATTT CTTACAATT ACTTGTAACA TCTAGTTCAC 11040
 TAACGGAAGA CTTTTATAAA TTAATTCGTG GTGAAGAAA AGCTAAAACG CACCAAAAAG 11100
 AATTTGTTAT GATTGGAAGA TTATCTGTAT TAGTTGTAGC AATTGTTGCC ATCGCGATTG 11160
 40 CATGGAATCC AAACGACACA ATTCTAAACT TAGTAGGTAA CGCTTGGGCC GGATTTGGTG 11220
 CATCGTTCAG TCCACTTG TG CTATTTGCAC TTTACTGGAA AGGTTTGACA CGTGCCGGTG 11280
 CTGTAAGTGG AATGGTTTCA GGTGCCTTAG TCGTTATCGT TTGGATTGCA TGGATTAAAC 11340
 45 CATTGGCACA TATCAACGAA ATATTCGGCT TATATGAAAT TATTCCTGGA TTTATTGTAA 11400
 GTGTAATCGT TACATATGTT GTAAGTAAAC TTAATAAAA ACCTGGTGCA TTTGTTGAAA 11460
 50 CTGACTTAAA CAAAGTTCGT GACATCGTTA GAGAAAAATA ATTCATAAGT CTTAACAAAT 11520
 TAAAAAGGTA CTAATGTAA TCAAAATTAT GACTAACATT GGTACCTTTT TATTATCTTT 11580

55

	AATTAAAGCA	CGTGGTTGGT	TACCATCTTT	AATACGAATT	TCATAGTTAT	CGATTTTATC	11700
	GAAATATTTA	TTGCGTTGTT	CAGTAACGTA	CTGTGTAATA	CCAATTGTTT	CAGCTTGTC	11760
5	ATAGTAATCG	ATTGGTAAAT	CTACTACTAA	TCGTTGTGGC	TTTTTATCAA	CAAATTTAAC	11820
	TTTCCCTACT	GCTTGTGTGA	AATTAGAAAA	ATATGATTGC	AAATTATCAT	TAAATTGCTT	11880
	GAAATTATTIA	TTTAAATTTT	CATCATAATC	TGCTGCTGTT	GAAGAAGGTA	ATAAAGCTGA	11940
10	TTTTTCATTG	ATATTATGCC	ATTCATTAAG	CTTTGTTTGA	CTCTTTTCTG	CAGTCGCTTG	12000
	AGTGATAAAT	TCACCTGGTG	TGATTGAATC	TTCACTTGAT	TGCTTATAAA	TTGCAAAATG	12060
	AATTGGTATA	TCTTTTAAAT	CATCATTTTC	ACGTAACCTT	GATAATATCT	CACTAGCCAT	12120
15	TTGTTTACCT	TGCTTTTTTAA	CTcGCTATCA	TCTAGTTTTT	TACTAAAAGT	CGATCCATCT	12180
	TTTTCTTTTT	TATAGTAATA	AACACTATTC	ATAGCTAAAC	CAATCGTCAT	ACCTTTAATA	12240
20	TTCTTACCTT	TTGTATCTCC	ACCACCATAA	AAATCTTGCT	CTAAAATGTT	AGATAAATAG	12300
	GCTGGTGATT	TTTCTGCAAT	CTTTTCAGGA	TCTGTTTCAC	CTcCGTGTGA	TGGATTAAGT	12360
	CCTAAATTTT	CATTGCGTTT	CTTGTCTTTT	TTATCTTTTT	CAGACATTTT	ATCGATTTC	12420
25	CGTTTTGTAT	ACTTAGGATT	TAAATAGGCA	TTAATTGTTT	TCTGTCCAA	AAATTGACCA	12480
	TCTTGATACA	AATATTTATC	TGTTGGAAAT	ACTTCTTTAC	TTAAGTTCAA	TAAACCATCT	12540
	TCAAAGTCGC	CGCCATTATA	ACTATTTGCC	ATGTTATCTT	GTAAAAGTCC	TCTTGCCTGG	12600
30	CTTTCTTTAA	ATGGTAACAA	TGTACGATAG	TTATCACCTT	GTACATTTTT	ATCCGTTGCA	12660
	ATTTCTTTTA	CTTGATTTGA	ACTATTGTGA	TGTTTTTGAT	TATCTTTTCC	AGCCTGGTCA	12720
	TCCTTATGGT	TACCACAAGC	AGCGAGTATA	AAGATAGCTG	TAATCAATAA	TACTAATGTA	12780
35	CGCTTCATCG	ACATACCCCT	CTAACTATTT	AAITCATTTT	GCTTATCTAC	AAATTGTTGC	12840
	TCTGTCCAAA	TTTCAATACC	TAAACTTTGT	GCTTTTGTTA	ATTTTGAACC	TGCATCTTCA	12900
40	CCAGCAATAA	CGACATCTGT	ATTTTTAGTA	ACGCTACTTG	TAACCTTAGC	ACCTTGATG	12960
	GCAAGCCATT	TAGATGCTTC	ATTGCGTGTC	ATTTGATGTA	GCTTACCAGT	CAGTACTATC	13020
	GTTTTACCAC	TAAATTCAGG	ATGTCCTTCA	ATATCTGATG	TTTTGATACC	TTTATAAATC	13080
45	ATATTAACAT	GTTTATCTTT	TAATTTTGA	ATTAAAGCAC	GAATATCTTC	ATTTTCTAAA	13140
	TAAGTAACTA	CAGATTGTGC	TACTTTATCA	CCTATATCAT	GAATTTCTAC	TAATTCCGCT	13200
	TCAGTTACCG	TTAGTAATCG	ATCTATCGTT	TCATATTTTT	CTGCTAACAC	TTGGCTCGCT	13260
50	TTAACACCTA	AATGCCTAAT	ACCTAGACCA	AATAATAAAT	TTTCTAAAGA	GTTGTCCTTA	13320
	GCTTGTGAA	TGGCAGCTAA	TAAATTATCA	ACTTTTTTCT	GCCCCATTCT	GTCTAAAGGT	13380

55

TAAAGCTGTT GAATAATTTT AGTGCCTAAA CCATCAATAT TcATGGCTTG TCTTGaTACA 13500
 AAGTGnATCa ATCctTcAAC AAGTTGTGCT TGGTCATTTT GG 13542

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1893 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

CAGTAAACAC CTCTGATTAC GAATATTTAT ACATTTATTT TAACACATGC ACTGATTTAC 60
 GACTACTAAA CACCTTTTACG TAAAAAGGGT AAACATGGTT TATCTATCTT GGTATCTAT 120
 TTATAAATAT TnTCATATT ACGCATAACA ATTGCTTAAA ATATGTATAA AAATGAATAT 180
 ATGTGTAATA AACCTGCTAA TTATTAGATT TAATAAGCGT CAATTGTTTG AACATATTcA 240
 ATTAAATCA CATTGATATC ACAGATACGA ATATTGTCGT ATAGAAATTG AAAATTCTAT 300
 TTTTAAATG AAAGTCTTCA ACATAATTTT AAGTTTCAAC ATGAGAAAAA TCGATTAACA 360
 AACACGTC A GTTGAATATG CCTTTTGAGA CATTTCAAAC TTTACAATTG TTGCTAATCG 420
 ATATATTTGC TTTTAGTGAT CCCTGCTATA AAATAAATCA ACGATTTCTA ATAAGTGTTC 480
 TGTATTGAAT TGTTTCATCAA TTTGCGTTAG TTCATCCACT GCTGCGTCTC TATGATAAGT 540
 CAATTATCT TCTGCGCCAT CTTTCCCTAA TAAACTCAG TACGTACTTT TATTATTTTC 600
 AAGATCGCTG CCCACTTTTT TACCTAACTT TGCTTCATCA CCATAGCAGT CTAATAAATC 660
 ATCTTTAATC TGGAAACATCA TACCTAAATG ATAACTATAA CTTTCTAAAT GTTCTTTAGT 720
 TGTATCATCG ACATTAGCGA TATCTGCTGC ACTCATAACC GCAAAAGTTA ATAATGCTCC 780
 TGTTTTTGTT TTGTGTATCA TTTCCAAAGT TTCAAGATCA ATTGGTTGGC CTTGCTTTG 840
 CATATCTAAC ATTTGACCGC CGACCATTC AACATGACCA CTTGCTATTG ACAGCCGTTG 900
 TAGAACTTTT ATTTTACTT CATCAGTTAA TCTATCATCA CTTGAAATAA GTTCAAATGC 960
 TTTAGTTAAT AAAGCATCAC CTGCTAATAT CGCAGTCCAC TCACCATATA CTTTATGATT 1020
 TGTTAATTTT CCTCGTCGAT AATCATCATT ATCCATCGCT GGTAGGTCAT CATGAATAAG 1080
 TGAATATGTA TGAATCATTT CTAGTGCAAT TGCGCTCTC ATACCTAACT CATACTCGGT 1140
 ATTTAGTGAA TCTAAAGTGA GTAATAACAG AACTGGTCGG ATGCGTTTAC CTCCAGCATT 1200
 TAATGAATAC AACATACTTT CTTCTAGCTG AGTATCCATT ACTGATTTAT TTATCGCAAC 1260

CATCCTCAGC TTCTTCTTTT ATTAAGTCAT TCACCTTTT TTCGGCATT TTTAAAGTTG 1380
 TGTACAAGC TGCTGATAGT TTCATACCAC GTTGATATAA ATCTAATGAT TCCTCTAAAG 1440
 5 ATACTGTTTC ATTATCTAAT TTTTGAACAA TTTGCTCTAA TTCTTGCATC ATTTCTTCAA 1500
 AACTTTGCGT TTCITTAGTC ATTATTACAC CTTACTTTCG TAACTTTTGC ATCTACTAAG 1560
 CCATCTTTCA TTGTTAACGT CAATTGATCA TTTTCTGTTA AATCTTTAGT ACTCGTAATG 1620
 10 ACTTCGTCTT TTTTATTAAAC AATTGCATAT CCACGCAACA TTGTATTAGT TGGACTTAAA 1680
 TTGTTTAAGT TTTCTACTTT ATTTTTCAAA TCATTTTAT AACTTAATAT CTTAGAATTC 1740
 AATAATTTAA CAAGTTGGTT TGTCAATTGA AGATTATnTT GTTGTTCTTG ATTAACACTA 1800
 15 CTTAGTAATG CTTTAAATn ATAACGTTGG TGCAACAGCA TTAAATCGAG GCCCCGGTGG 1860
 TCCAAAGTTG CCCGAATTnG TGGTTTCAGG CCC 1893

20 (2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 821 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

30 AAAATATATT CCTTCACTTA ATATTCAATT AGAGAAAAAC ATGGTGATTG TAATATGTTG 60
 TGCAATATTT CTGGGTGTTT TAATACTTTT TTTATTCTG AATCGTAAGC TAAGGTTGGA 120
 AATTTATAAT AATAACTCTA GTAAAGGGAA AATAATTTTA TTTCCTTCAT TAAAAAATT 180
 35 TTGTTTCACA ATATTTTATT ATTTTTTATT TGGCGGTCTT TCAATAATGG CTCTAAGTAT 240
 GTTATTAAC TTAATCCTC AAAATATAAT AGGCTTTATT GGTGTTGG TAATGACTGC 300
 AGGTTTCTTT CTGTAAACA TGTCATCGAT TATTGACAAA AAAATTTATG TATTATCTAA 360
 40 AACTAACACG GTGAAAAAT GATGGTTTAG CTGGATTTAC TGCAGGTTCT ATTTCCGCAA 420
 TACTTGTATA TTGGACCAAT CAAAAAATG AATTTGGAAT AAAAGATAAA AACGATTGGA 480
 45 TAGGACATAA ACTAGACGTT GGTATAGATG CTGTAGAAAA ATCTGCAGAA AAAACAGTAG 540
 ATGGTGTTGA AAATGTCATG GTGAAGCTTC AAAAGTATT TCTAATCATA TAAGCCCTAA 600
 GAAATGGAGC TGGTAAATGT TGCTATGCGA ATCTAAAATC ATCAATAAAA ACCCAAATA 660
 50 TAGAATTATT AAATATAATG ATGAATACTT AATGGTCGAT ATAATAAGCA CTTGGATTAG 720
 TTTATTTTTT CCTTTTATTA ATTGGTTCAT CCCaAAAGaA TACGTCAAAA TTAGTAGAGA 780

55

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2343 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

```

AGTAAGATAA TTTTCAATTA GAAAATATCT TACTGCTGTT CTCTATTTAT ACAATACTTC      60
GTATTGAATG GcTTTCGCTTT CCTAGGGTGC CGTCTCAGCC TTGGTCTTCG ACTGGCACTG      120
CTCCCTCAGG AGTCTCGCCA TTAATACTAC GTATTACAT GTAATTTTAC TTTGAAATAC      180
TTTTAAAAAA TAAGACACTT TGCCCAACTT GCACATAAAT GTAAAATTCA ATAAAATGAA      240
TTTTCTGTGT TGGGTCCCTT CTTATAATTT AATAAATACC ACTAACTAA ATTAACGAGG      300
TGCCTTATGT ATAAAAATTA TAACATGCCC CAACTACACT ACCAATAGAA ACTTCTGTTA      360
GAATCCCTCA AAATGATATT TCACGATATG TTAATGAAAT TGTTGAAACr ATACCTGATA      420
GCGAATTCGA TGAATTCAGA CATCATCGTG GCGCAACATC CTATCATCCA AAAATGATGT      480
TAAAAATCAT CTTATATGCA TATACTCAAT CTGTTTAATT ATGTTCAAAG CATTAAAGTA      540
ACAAGACAAT ATCTAAGATA TCAAAGATAG AAATTTTTTG ACGTTGTTGC TGATTGTAAA      600
CATAACCATC AATTCATAA TTAATAGCAT CAATACGATA AATGGTTAAG CGTACTGAAT      660
CTACAAAGCC ATTATTATAA AATTTAACTT CTACAGGTTG GGCATATTGT AGCGCCTCGT      720
GTAGCCGAAT GTTTAGCTCA GCCAATTGAT CATCTGATAA TACAGGACGT GTAATTTTGT      780
TTTGGTCGAT AATGTATTGT TGAATCGTTT CGAATTGTTT GGGTAATGTT GCAAAAGGAG      840
CCCA1TTTAAT CATGCCTCTT CCCATAGGTA TATTGTTATC TAGTAATTCT CTTGGAACGT      900
TACGATAATC AGTTTCTTCT TCATAACTTG TCATCCTTAA TTCACCCCAA TCTGATAATT      960
ACATTATACG AACATGTGTT CTATTTTGCA ACAAAAATTT TGTGGaAGCA TAAACGCGTT     1020
AATAATTAAT GCTCGTGtAA GTAAAAAAGA GGGATTAATT AAAATCGAAT AATGaCATAT     1080
CACaGCAAAT AGTTCTTTTA AAGTAGTTAA ATAGTTTTAG CTTTAAGGAA aTGATAaaTG     1140
ATTGTwaATT CTAGCTAAAA TTTAATAAAA TGAAAATAAG ACTAACATGG AGGGGTAAAA     1200
GTAATGACAA ATGGATATAT TGGTTCTTAC ACTAAAAAGA ATGGTAAAGG GATTTATCGT     1260
TTTGAATTAA ACGAAAATCA GTCACGTATT GATTTATTAG AAACAGGATT TGAATTAGAA     1320
GCGTCTACAT ATTTGGTGCG TAATAATGAA GTTTTATATG GAATCAACAA AGAAGGAGAA     1380

```


TGT TTGTCTT CAAAAGCTGG TACAGGTTGT TATGTATCGA TTTCAGAAGA TAAACGATAT 1500
 TTATTTGAAG CGGTATATGG TGCTGGCATC ATACGTATGT ATGAATTAAA TACGCACACA 1560
 5 GGTGAAATTA TACGTCTAAT TCAAGAACTT GCACATGATT TTCCAACAGG TACACATGAA 1620
 AGACAAGATC ATCCACACGC ACATTATATT AATCAAATC CAGATGGTAA GTACGTTGCA 1680
 GTAACAGATT TAGGTGCTGA TCGTATCGTT ACTTATAAAT TTGATGACAA CGGGTTTGAA 1740
 10 TTTTATAAAG AATCTTTATT TAAAGATAGT GATGGGACAA GACATATTGA ATTTTCATGAT 1800
 AATGGAAAAT TTGCTTATGT CGTACACGAA TTATCAAATA CTGTGAGTGT TGCAGAATAT 1860
 AATGACGGTA AATTTGAAGA GCTCGAGCGT CATTTAACAA TTCCTGAAAA CTTTGATGGA 1920
 15 GATACTAAAC TTGcAGCAGT GCGTTTATCT CATGaTCAAC AATTCTTATA TGTATCTAAT 1980
 AGAGGGCATG ATAGCATTGC AATTTTAAAA GTTCTTGATA ATGGTCAACA CTTAGAACTA 2040
 20 GTAACAaTTA CTGAAaGTGG TGGTCAATTC CCAAGAGATT TTAATATTGC CTCATCAGAT 2100
 GACCyTTTAG TTTgTGCTCA kGaGCaAGGA GATTcAGTTG TAACTGTTTT CGAAAGAAAT 2160
 AAAGAAACAG GTAAAATTAC GCTATGTGAT AACACTCGTG TAGCATCTGA AGGTGTATGT 2220
 25 GTCATATTTT AATCTTTAAT TAATCATGAT AAAAAGAAAA CCATGTTTCC AAAAAATTTG 2280
 TGTATACCTT GAAATTTATT GnTTTCCAgN ACATCAATTA TGGGAAGCAT GGnTTATTTT 2340
 TGT 2343

30 (2) INFORMATION FOR SEQ ID NO: 158:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4837 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

40

AAATTGCCAG TTGGTATCGC TTCTGGTGCA GTAGTCGAAG GTTCTTTCCA AGGTATCATT 60
 CCGATTGGCT ATATCGTTAT GATGGCAGTA TTGTTATACA AAATTACTGT TGAATCTGGA 120
 45 CAATTTTTAA CAATTCAGA TAGTATTACA AATATTTAC AAGACCAACG TATTCAAGTT 180
 TTACTTATTG GATTTGCATT CAACGCATTT TTAGAAGGTG CAGCAGGATT TGGTGTACCA 240
 ATTGCAATTT GTGCACTTTT ATTAAACACAA TTAGGATTTA ATCCATTAAA AGCTGCGATG 300
 50 TTATGTTTTAG TCGCAAATGC AGCGTCTGGT GCTTTTGGTG CGATTGGTAT CCCTGTAGGT 360
 GTTGTAGAAA CGTTGAAATT ACCTGGAGAT GTTTCAGTAT TAGGTGTTTC TCAATCAGCA 420

55

	GGTTTTAGAG GTGTAAAGA AACATTACCA GCAATTTTAG TAGTTTCAAT CACTTATACA	540
	CTTACTCAAG GATTATTAAC TGTATTAGT GGACCTGAAT TAGCAGATAT TATTCCACCG	600
5	TTATTAACAA TGTTAGCATT AGCAGTATTT TCTAAAAAAT TCCAACCAAA ACACATTTAT	660
	CGTGTTAATA AAGATGAAGA AATTGAACCT GCAAAAAGCAC ATTCTGCAAA AGCAGTATTA	720
	CATGCATGGA GCCCATTCAT TGTATTAACA GTCATTGTAA TGATTGGAG TGCGCCATTC	780
10	TTTAAAACT TATTCTTACC AAATGGTGCT TTATCATCAT TAGTATTTAA ATTCAACTTA	840
	CCTGGaACAA TCAGCGAAGT TACGCATAAA CCATTAGTAT TGACTTTAAA TATTATTGGA	900
	CRAACAGGTA CAGCTATTTT ATTAECTATT ATTATTACAA TTTTAATGTC TAAAAAGGTT	960
15	AACITTTAAAG ATGCAGGTAG ATTATTGCGC GTTACATTTA AAGAGTTGTG GTTACCAGTT	1020
	CITACAATTT GTTTCATCTT AGCAATTTCT AAAATCACAA CTTATGGTGG TTTAAGTGCA	1080
20	GCAATGGGTC AAGGTATTGC TAAAGCAGGT AATGTCTTCC CAGTCTATC ACCAATTTTA	1140
	GGTTGGATAG GTGTGTTTAT GACAGGATCA GTTGTAATA ACAACTCATT ATTTGCACCA	1200
	ATTCAAGCTT CTGTTGCACA ACAATTGGA ACAAGTGGT CACTTCTTGT ATCTGCTAAT	1260
25	ACAGTTGGTG GTGTAGCGGC AAAATTGATT TCACCACAAT CAATTGCAAT TGCAACTGCA	1320
	GCAGTAAAC AAGTTGGTAA GGAATCAGAA TTATTAATAA TGACATTGAA ATACAGTGTA	1380
	TGTTTACTAA TATTCATCTG TATTTGGACT TTCATCTTGT CATTATTATA AAAAAACGTA	1440
30	TTTCAAAATA TAAATATACA GAAGGTGAGA TGTTTTCTAA CATCTCATCT TTTTTTATG	1500
	GATCATTAAAT GAAAGAAGTT TGACATTATA ATAATGGTAG CGCTTTATGT TAAAATGAAT	1560
	AGTGAGTAAT CAGCAATCAA ATTAAATTGG TTGATAGCTG TTAAGGTTTG TGGTTTTGTC	1620
35	TTTGTGCTAT CGCnCATAAA GTATATAATT AAAGTAGTTT CGTTATTATA AAATATTAAT	1680
	ATACATAGTA GATAGTAATA GAGCATCACC ATGGGAACCT ATTGAGACAC TTATTGATTT	1740
40	AAAGTGGTAT TAATATGTCG TATTTCTCGA ACGTTCCATT ATTCATTTTA AAAAGGGGGA	1800
	CTGTATTTGT TATGACAACA CAACATAGCA AAACAGATGT CATCTTAATT GGTGGCGGTA	1860
	TTATGAGTGC aCATTAGGAA CATTACTTAA AGAATTATCA CCTGAGAAAA ATATTAAAGT	1920
45	GTTTGAAAAA TTAGCACAACT CTGGCGAAGA GAGTTCAAAT GTATGGAATA ATGCCGGTAC	1980
	AGGGCATTCA GCACTTTGCG AGTTGAACTA TACAAAAGAA GGTAAGGATG GCACAGTTGA	2040
	TTGTAGTAAA GCAATTAAGA TAAATGAGCA GTACCAAATT TCAAAACAGT TTTGGGCATA	2100
50	TTTAGTTAAA ACAGGACAAT TAGATAACCC AGATCGCTTT ATTCAAGCGG TGCCACACAT	2160
	GAGTTTTGTC ATTGGCGAAG ATAATGTAGC TTTTATAAAA AGTCGTGTTG CAACGTTAAA	2220

55

	GGTACCGTTA ATGATTGAAG GTCGTAAGTC TGATGAACCA ATTGCTTTAA CTTATGATGA	2340
	AACTGGTACa gATGTTAACT TTGGTGCGTT AACTGCAAAG TTATTTGATA ATTTAGAGCA	2400
5	ACGTGGTGTG GGAATTCAAT ATAAGCAGAA TGTATTAGAC ATCAAGAAAC AGAAATCTGG	2460
	GGTATGGCTA GTTAAAGTTA AAGATTTAGA AACTAATGAA ACGACAACAT ATGAATCTGA	2520
	TTTTGTATTT ATTGGTGCTG GCGGTGCGAG TTTACCATTA CTCCAAAAGA CTGGGATTAA	2580
10	ACAATCAAAA CATATTGGTG GTTTCCCGGT AAGTGGATTA TTCCTGCGCT GTACAAATCA	2640
	AGAAGTGATT GATCGTCATC ATGCTAAAGT GTACGGAAAA GCAGCAGTGG GTGCGCCACC	2700
	AATGTCAGTG CCGCACTTAG ATACACGTTT TGTAGACGGC AAGCGTTCAT TGTATTTTGG	2760
15	TCCATTTGCA GGTCTCTCAC CTAAATTTTT AAAACAGGT TCACATATGG ATTTAATTAA	2820
	ATCGGTAAAA CCAAATAATA TCGTGACGAT GTTATCTGCA GGTATCAAAG AAATGAGTCT	2880
20	TACGAAGTAT TTAGTGTCAC AATTGATGTT ATCTAATGAT GAGCGTATGG ATGATTTAAG	2940
	AGTCTTTTTTC CCAATGCTA AAAATGAAGA TTGGGAAGTG ATTACAGCAG GGCAACGTGT	3000
	CCAAGTAATC AAGGATACTG AGGATTCTAA AGGTAACTTA CAATTTGGTA CTGAAGTTAT	3060
25	TACGTCAGAT GATGGCACAT TAGCTGCATT ACTTGGTGCA TCACCTGGTG CGTCAACAGC	3120
	TGTAGATATT ATGTTTGATG TTTTACAGAG ATGCTATCGT GATGAATTCA AAGGATGGGA	3180
	ACCAAAGATT AAAGAAATGG TGCCGTCATT TGGTTATCGC tTAACAGATC ATGAGGATTT	3240
30	ATATCATAAA ATTAATGAAG AAGTAACTAA GTATTTACAA GTTAAATAAT AAACGAAACG	3300
	GTAATGTCTT TTTAATGTG ATAGACATTA CCGTTTTTTA GTGGTTAATA AAAATCATT	3360
	TAATTGTTTC AGTTGCTTGT TAATAGTGTC TACGTAGTTC TTGTTTTTAA AGAATTGAAT	3420
35	TATCCAAATT AATACATAAA CCACAATGAA GATAATTGTG AATATGATTA GATAATGCAC	3480
	TGTTAGTGGA AACCAACCGG CAAGCATTGC TAAAGGCAAG AATCCGACAT ACGTTGTTAT	3540
	GAAATGCATT ATAGTTGCTT TAGTAATGCT CCAATCTGTG TATTTAAAGA TAAAATCTCC	3600
40	AAGGAAAAAG ACGACGCCTA TGAGTAACCA TAAAATGATA GAAATCAACA TTACGGTAGT	3660
	TTCTGTGAAA TCGGTATAAT ACAATATGCC AATAGTTGAT TGTGGGTTCa GTGGATAATA	3720
	TTTGCCGTCT GCAAATAACA TACTAAAGAA CAGTGAAAGG GACAAACCAA TGATTAAGCT	3780
45	AATAAATAAT GAGTTTTTCA AATTTTTCAT ATTGATAAGC GCTCCTTTAT AGATTTTAAA	3840
	TAACGTCTAG AAGAATAGGT GTAGTGTCa TCTTTAAGAT ACATACGTAT AAGTCCATTT	3900
50	GGCTCTAATA ATAATTTTTC AATGTAATAC TTGTTGACGA TTTCTGATTT GGAAATGCGA	3960
	ATGAAATGTT GTGGTAACTG TTTTCTAGT TCATAAAGTC GTAATTTTAG TTTGAATTTT	4020

55

ACATTAATGA TATGGATTTT TTTGTCTATG TATCCGACTA ATGTATGTGA TTTGTCTAAA 4140
 TCATTGACTG CATTAAATAAT ACTTTGAACG TTATCATTCA TTTTAGGTGC ATGTATATCA 4200
 5 ATATAAGATT CCGTCTCATT TGCATTGATA AATAAATTGA GTTTCATCAT AGGTTAATGC 4260
 CTCCTTCAAA ATTATTAAAC CATAAATGAC CATCGATATA TTTAAATTTT GTTGAATGGT 4320
 AGAAATTAAA TGTTAAGTGG CTAGAAAGCG CTAATCAATA TAAAAGATAC CTCCTGAAAT 4380
 10 AAAAAACAGAA ATGTTTTTTC AGGAGGTAGA GATTAAAGTG AATTATTTGG CAGTGTAATA 4440
 GTAAAGGTGG TTACATACTC GTTACTTTGT GTGAATTGGA TTGTACCATG ATGCAATTCA 4500
 ATGATGGATT TTGTAATTGC AAGACCTAAA CCATTGCTAT TATCATGTTT GCTCACTTTA 4560
 15 TAAAAACGTT CAAATAAACG TGCTTCAGCT TGTGGACTAA TTGGTGAACC ATCATTACTT 4620
 ATTGTGAAAA TGATATTGTT GTGACTATGT TGCAAAGCGA TGTCAATGGC ACCACCAACA 4680
 20 TCTGTATACT TAATAGCATT TATTAATAAA TTA CTCAATG CTTGATGTAA CAAACGTTGA 4740
 TTTCTAGGA AATTGATGAT TCTAGGTCAG CTAAnATGAT TAACGACTTT TCATCAGCAG 4800
 CAnATTGTTT ATGTCGAATG ATATCnTTAA TGAGCTG 4837

25 (2) INFORMATION FOR SEQ ID NO: 159:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1600 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

35 ACAATTATTG GATTATTATC AAGCAACGTT AATGGATGAC TTCCACTTAC AACAGAAATG 60
 CCCATAGATT CTAAATCTtT TGCATGAGCA TCTTGTGATA AGTCTTTTCC ATCATTGACA 120
 40 GTTACATTCT CACCTAATTT ACTTAATAAT TTAGCTGCTT CATAACCACT TTTTGCCAAA 180
 CCGACAACCTA ATACATTTTT ATTTTCTAAC CCTGTATAAT TAAGCATCTT AATGCACTCC 240
 AATCCATAAA CCGATTAAAC CTGAAATCAG ACCAACAGCC CAAAATACTG TAACTACTTT 300
 45 CCATTGCTC CATCCTATCA ATTCAAAATG ATGATGAATC GGACTCATT TAAATATACG 360
 CTTTCCAGTC AATTTAAAGC TAGCGACTTG TAACATAACA GATAATGTTT CAATTACGAA 420
 TACTAAACCT ATAAAAATTA ATGATAATTC CTGATTAAAGC ATGATTGAAA TGGTAGCAAA 480
 50 TATACCACCT AAAGCTAAGC TACCTGTATC TCCCATAAAC ACTTTAGCAG GGTTAATGTT 540
 ATATGGTAAA AATCCTAAAA GTGCAAACAA CATAATGATA CAGAAAATAC CAATTGCCGT 600

55

TGCTAATCCA TCTAAACCAT CTGTTAAATT TACTGCATTA GAAAAACCTA CTTGCCAAAA 720
 AACAAATGAAA ATAACATATG CAAATGATAG TGGGATTGCT ACATTGCTAA ATGGAATATG 780
 5 TATGCTCGTA GAAAAATTCA CCAAATGAAA CACATTACTT AAAACAAAGA ATATAATCGC 840
 AATACCAATT TGCGCCAAAA ACTTCTGTTT ACTTGTTAAA CCTTGGTTAT TCTTTTTAAC 900
 AACAAATAATA TAATCATCTA TAAAACCAAT TAACCCAAAA CCAATCGTCA CAAATAATAA 960
 10 CAGTATGATT GGATTAGCTT GATCTACAAA TATAATAGCC ACCAAAGACG TTATCACAAT 1020
 ACTTAATAGA AATGTTAGTC CACCCATCGT TGGTGTACCA GTCTTCTTCA TATGGCTTTG 1080
 TGGACCTTCT TCTCGAATAC TTTGACCAAA TTTTCATCCTT TTTAATGTAG GTATTAAAC 1140
 15 AGGTACCAAA ACAAATGTAA TCACTAGCGC TAATAACGCA TATACAAAAA TCATAACTAT 1200
 CTCCTCTTCT TAATCCAGAC TTTTTTAACC ACTAATATAT TATCAATTTT TCAATTAAAT 1260
 20 AAACAAAGTT GTAATCAAAA TTTATAATTT TTCTTTTITA CGGCATAAGA GGCCAGTATA 1320
 AAAAGTTTGC CTATAACAAA CAAGTTAATC TGACCTCGTC TACCTTAAAA TTCTCTATCA 1380
 ACACCTATTT ATAAAGATTA AATGAAGATG TTGTTTTCTA TCACAGCATT ACTTTAGTAA 1440
 25 AAACAAATAG TGACAATACA TCCTAATTTA ATGTAGCCAT TCTTGTTAGT CCGACTTATC 1500
 CTTGTCAGTT TTAATGTCAG ATTTCTTCTT ATCATCTGAA TTTGAATCAG AATTATTCGT 1560
 CGAATTGCTG TCTACATTCT CTGGATGGAA AATTCTACGT 1600

(2) INFORMATION FOR SEQ ID NO: 160:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1186 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

ATTGCCTTTG TTTTAATTTT AAATCAAAAT mGCCTATGAA AGATTTAAAT CAATTAATTT 60
 CTATAATATT ATCATTTTTA AAGCATATCA TTGTTTAGTT TTTTATAAT TGGATAAATA 120
 45 CTAATAGTTA CTTTATAAAA CATTACATAG AGAAAGGTTA AGGAGTGCAC ATGTCGAAAA 180
 AGGATCACTC TTCTTCAAAA TACCTTAATT CTGTTAAGGA AGCGCAAGAG GAGTCAAAAA 240
 AGAAAAATAA AAGTAATCCC AAAATTGATG TTGATCGTAC ATATATTGAA CCTCAACAAT 300
 50 TCCAATCTAA GAAACCTAAA AAAGATGATC AGGTTTTCTT CTTATCAAGA TTAAATAAAC 360
 CTGCAAAATA TAAGAAAGAC TCTAATTTCT TATCATATCT CATCTATCGC ATAGGAAAAG 420

TGTGCTTTT CCTATTAACA TTATTACCAT TTTTCAATAT TAAGCAGAGT CAAATTACTA 540
 ATATGTTAAG CAATGCACCC GCTGAAACAT CTACTCTAAT TAAGAGTGTA ATTGGTGATA 600
 5 TAACTCAAAA CTCCAGTGGT GGCTTATTAT CTATCGGTTT GATTTTAGCA ATTTGGTCAG 660
 CTTCAAATGG AATGACTGCA ATTATGAATT CTTTCAATGT TGCTTACGAT GTAGAAGATA 720
 GCCGTAATGG AATCGTATTA AAATACTAA GTGTGTCTT CACTGTAGTT ATGGGCGTTG 780
 10 TGTTTGAGT TGCTCTAGCA TTACCAACGC TTGGTTCTGT AATTAGTCAT TTCCTATTCG 840
 GTCCACTTGG ATTTGACGAA CAAGTGAAAT GGATTTTAA CCTATTAGA ATTGTGTTAC 900
 CAATCATTAT TATATTTATC ATATTTATCG TGTTATATTC GGTGACCT AACGTTAAAA 960
 15 CGAAGCTTAA GTCAGTATTA CCAGGTGCAG TATTTACTTC AATTATTTGG TTAGCTGGTT 1020
 CATTTGGTTT TGGTTGGTAT ATTTCAAATT TTGGTAACTA TTCTAAAACA TATGGCAGTA 1080
 20 TCGCGGGTAT CATCATTTTG TTAATATGGT TATATATCAC AAGTTTTATT ATAATTGTCG 1140
 GAGCTGAAAT CAATGCAATC ATTCATCAGC GTAGTGTAAT TAAAGG 1186

(2) INFORMATION FOR SEQ ID NO: 161:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7872 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

TCTTGAGCCA TCTTTTGAGC TAACTGACTA GATTGATACC CAAAAATCAT AGTTACCAAC 60
 35 ATAAACTTTA ATTTTACCGA AGTCTAAATC AGCGATATGA GTACATACAT TATTTAAGAA 120
 ATGATCGGTCA TGCGATACTA CGATAACAGT ATTATCAAAG TTAATTAAGA AATCTTCTAA 180
 40 CCAACTGATT GCTGGAATAT CGAGACCGTT AGTAGGCTCA TCCAGTAATA GTACGTCTGG 240
 TTCACCGAAT AAACCTTGCG CTAATAATAC TTAAATTTTT TGGTTGTTTT CTAATTCAGC 300
 CATTTTTTTA TCGTGTAAG TTGGATCGAT ACCTAAACCA GATAAAAGGT TAGCAGCATC 360
 45 AGCTTCAGCA TTCCAACCAT TCATTTCTGC AAATTCACCT TCAAGTTCAG CAGCACGGAT 420
 ACCATCTTCA TCACTGAAAT CTGGCTTCAT ATAGATTTCA TCTTTTTCTT TCATAACCTC 480
 ATAAAGACGT TCGTGACCTT TAATTACAAC ATCAAGCAGC CGTTCATCTT CATAAGCATA 540
 50 GTGGTCCTGT TTTAAACAG CTAGACGTTT ATTTTCCCT AATGAAACAT GTCCTGTTTG 600
 AGAATCTAAT TCACCAGATA ATATTTTAA GAATGTTGAT TTACCTGCAC CATTGCGACC 660

55

	ATCTCCAAAA CGTAAACTCA CATCAGTTAC TTGTAACATG CATTTTCTCC TTTTTCAT	780
	TCGATATTCT AACGGAAGAA TTATATCATA TTATCGTCAC AGTTTCGACC TCATATAAGT	840
5	TGTAATGATA GAATGACTCA CACATGTTAT AATAATAAAG AATACAAGAA TCGAAGGAGA	900
	ATAACATGGC ATTAGACAAA GATATAGTAG GTTCTATAGA ATTCCTTGAA GTAGTAGGGT	960
	TACAAGGTTC AACTTACCTT TTAAGGAC CAAACGGTGA AAACGTAAAG TTAAACCAAT	1020
10	CAGAAATGAA CGATGATGAT GAATTAGAAG TAGGTGAAGA ATATAGTTTC TTCATTTATC	1080
	CAAACCGTTC AGGTGAATTA TTTGCAACTC AAAATATGCC TGATATTACG AAAGATAAAT	1140
	ATGACTTTGC TAAAGTACTT AAAACGGATC GCGATGGGGC ACGTATAGAT GTTGGATTAC	1200
15	CCCGTGAAGT GTTAGTACCA TGGGAAGATT TACCAAAGT GAAATCACTA TGGCCACAAC	1260
	CTGGTGATTA TTTGCTAGTT ACATTACGAA TTGACCGTGA GAATCATATG TATGGACGTT	1320
20	TAGCGAGTGA ATCTGTTGTA GAAATATGT TTACACCTGT ACACGACGAT AATTTAAAAA	1380
	ACGAAGTCAT TGAAGCCAAA CCTTACCGCG TATTACGAAT TGGTAGCTTT TTATTAAGCG	1440
	AATCAGGTGA CAAAATTTTC GTACATGAAT CAGAACGTAA AGCTGAACCA AGATTAGGTG	1500
25	AATCTGTTCA AGTTAGAATT ATCGGGCATA ATGATAAAGG TGAGTTAAAT GGTTCATTTT	1560
	TACCACTTGC ACATGAACGT TTAGACGATG ACGGCCAAGT CATCTTTGAT TTACTAGTTG	1620
	AATATGATGG TGAATTACCA TTCTGGGACA AATCAAGCCC TGAAGCGATT AAAGAAGTAT	1680
30	TCAATATGAG TAAAGGTTCA TTCAAACGTG CAATCGGTCA CTTATATAAA CAGAAGATTA	1740
	TTAATATAGA AACAGGTAAA ATCGCTTTAA CTAAAAAGG TTGGAGTCGA ATGGACTCAA	1800
	AAGAATAATC ATTTTACAC GTGTCGTAGG ATGCGTGTTT TTTTATTCA ATATTAAATC	1860
35	GGACAGATGA AGTAGTTTTT TAAACATTCC TTTCAAAGTA AAAAATTAAA TAATTCAAAC	1920
	GAATAGGCTG GGAATTAAAG TTCTTAGGCA ATGTAAAAA GCTGATTTCT ATTAATTATT	1980
40	TGATGGAAAT CAGCTTTTTT GATATGTATT TTATAATGTA CAGCTCGTTG AGCTGCTATT	2040
	TTCTTTATAT TAAGTGCCAT TAATACAAAA CCTAGCTCTC GTTTAACTTT ATTTATTCCT	2100
	CGAACTGACA TTCGAGTGAA ACCCAAAATA GCCTTCATAA ATCCAAAAAC AGGCTCTACA	2160
45	TAAATTTTTT TATGACTATA GATTTTTTTC GTTCTGGTT CAGAAAGCTT TTGAATTAAT	2220
	TGGGCTTTAA TGTATTTCAA AGTAAATTA CATGTTAATA CGTAGTATTA ATGGCGAGAC	2280
	TCCTGAGGGA GCAGTGCCAG TCGAAGACAG GGGCCCCAAC ACAGAAGCTG ACATATAGTC	2340
50	AGCTTACAAC AATGTGCCGG TTGGGGTGGC TGAGACGGCA CCCTAGGAAG GGACCCGTCA	2400
	TCAAAATTC TATTTATAGA ATTTTACAGT AATGTGACAG ACGGGCAAAG CGAAGCCATT	2460

55

	CTTACTGCTG TTTTTTTAGG GATTTATGTC CCAGCCATTT TTGTATTCAT ATTTAAATTT	2580
	CGATAATTTT TCAGGAAGCA TTTTAATTTT ACTAATGAAG CAATATTTT TAGATTAACA	2640
5	AAAATTAATA TTTACATTTT CTTAACAATT TTTTATGTAA CATTTACAGT TTCTAAAAAT	2700
	GAGGTTAATA ATTCAAGGTT AAGATAAAGA TGTAATCAAT ACAAATACTA TTTGTTGTTT	2760
	ATACAGGGAG GATATTTCAA TGAAAAAATG GCAATTTGTT GGTACTACAG CTTTAGGTGC	2820
10	AACACTATTA TTAGGTGCTT GTGGTGGCGG TAATGGTGGC AGTGGTAATA GTGATTTAAA	2880
	AGGGGAAGCT AAAGGTGATG GCTCATCAAC AGTAGCACCA ATTGTGGAGA AATTAAATGA	2940
	AAAATGGGCT CAAGATCACT CGGATGCTAA AATCTCAGCA GGACAAGCTG GTACAGGTGC	3000
15	TGGTTTCCAA AAATTCATTG CAGGAGATAT CGACTTCGCT GATGCTTCTA GACCAATTAA	3060
	AGATGAAGAG AAGCAAAAAT TACAAGATAA GAATATCAAA TACAAAGAAT TCAAAATTGC	3120
20	GCAAGATGGT GTAACGGTTG CTGTAAATAA AGAAAATGAT TTTGTAGATG AATTAGACAA	3180
	ACAGCAATTA AAAGCAATTT ATTCTGGAAA AGCTAAAACA TGGAAAGATG TTAATAGTAA	3240
	ATGGCCAGAT AAAAAAATAA ATGCTGTATC ACCAAACTCA AGTCATGGTA CTTATGACTT	3300
25	CTTTGAAAAT GAAGTAATGA ATAAAGAAGA TATTAAAGCA GAAAAAATG CTGATACAAA	3360
	TGCTATCGTT TCTTCTGTAA CGAAAAACAA AGAGGGAATC GGATACTTTG GATATAACTT	3420
	CTACGTACAA AATAAGATA AATTAAAAGA AGTTAAAATC AAAGATGAAA ATGGTAAAGC	3480
30	AACAGAGCCT ACGAAAAAAA CAATTCAAGA TAACTCTTAT GCATTAGTA GACCATTATT	3540
	CATTTATGTA AATGAAAAAG CATTGAAAGA TAATAAGTA ATGTCAGAA TTAATCAAAT	3600
	CGTCTTAGAA GATAAAGGTA AAGCAGCTGA AGAAGCTGGA TATGTAGCAG CACCAGAGAA	3660
35	AACATACAAA TCACAATTAG ATGATTTAAA AGCATTTATT GATAAAAATC AAAAATCAGA	3720
	CGACAAGAAA TCTGATGATA AAAAGTCTGA AGACAAAAAA TAATAAGACG CAATTTCAA	3780
40	TGTGTCTTGA AACATGATTT TGATGGTGAA TCATTATTTA GAGTACAAAG CTTGATTTAT	3840
	CGAGACGCTG ATTTTGACAT TCAGTTAGTC TACAAGCTTA TCAACTTAAA ATAGTGGTTC	3900
	ATCATTATTT TACAAATCTA ATTATTTTGG GAGTAATAGA AAGAGGTTTG ATTATGACTT	3960
45	CATCTACTAA TGTAAAGCT TTAATCGAAA AAAATAATAA TAAAAAGGA AAGCATAATG	4020
	ACAAAATTAT ACCAGTTATT TTAGCCGCAA TTTCAGCGAT TTCCATTTTA ACAACACTAG	4080
	GTATATTAAT CACATTGCTT TTAGAAACCA TCACTTTTTT CACCAGAATT CCAATAACTG	4140
50	AATTTCTATT TTCTACTACT TGGAACTCTA CCGGTTTCTA CCTAAGTTT GGTATCTGGG	4200
	CATTGATAAT AGGGACTTTA AAAATCACAG TTATTGCGAC TATATTTGCA GTTCCAGTCG	4260

55

	AACCGATATT AGAAATTTTA GCAGGAATAC CAACAATTGT GTTTGGTTTC TTTGCATTAA	4380
	CCTTTGTTAC ACCAGTATTA AGATCTTTCA TACCAGGTCT TGGAGAGTTT AATGCTATAA	4440
5	GTCCCGGCTT AGTTGTCGGT ATTATGATTG TCCCTCTCAT CACAAGTTTG AGTGAGGATG	4500
	CAATGGCATC TGTACCAAAT AAAATTTCGAG AAGGTGCCTA TGGACTTGGA GCAACTAAAT	4560
	TAGAAGTAGC AACTAAAGTC GTACTTCCCG CAGCAACATC AGGTATTGTA GCTTCAATCG	4620
10	TTCTCGCGAT TTCAAGAGCA ATTGGAGAAA CGATGATTGT ATCATTAGCG GCAGGTAGTT	4680
	CGCCAACAGC TTCATTAAGT TTAACAAGTT CGATTCAAAC AATGACTGGA TATATTGTTG	4740
	AGATAGCGAC AGGTGATGCA ACATTGGAT CAAATATTTA TTACAGTATT TATGCTGTAG	4800
15	GGTTCACACT ATTTATCTTT ACCTTAATCA TGAATTTACT TTCTCAGTGG ATTTCTAAGC	4860
	GTTTTAGGGA GGAGTATTAA TATGGAAACG ACAGATAATA ATAGACAATC ACTCGTCGAT	4920
20	CAACAACCTG TCCAAAAACA TTTATCATCC AGAACGGTTA AAAATAAAGT GTTCAAATC	4980
	ATATTTTTAG CATGTACATT ATTAGGACTT GTCGTACTTA TTGCGTTGTT AACTCAAACA	5040
	TTGATTAAAG GGGTAAGTCA TTTAAATTTA CAGTTTTTCA CTAATTTTTTCT TCTTCAACA	5100
25	CCATCTATGG CTGGCGTTAA AGGCGCGTTA ATCGGTTTAC TTTGGTTAAT GTTAAGTATC	5160
	ATTCCATTAT CAATCATCCT AGGAATAGGT ACAGCTATAT ACTTAGAAGA ATATGCGAAA	5220
	AACAACAAAT TTAATCAGTT TGTAAAAATC AGTATTTCCA ATTTAGCTGG TGTACCATCA	5280
30	GTTGTATTTG GGTATTAGG TTATACTTTG TTCGTTGGTG GTGCAGGGAT TGAAGCCTTG	5340
	AAAATGGGTA ACAGTATATT GGCAGCAGCG CTAACAATGA CCTTACTGAT ATTACCAATT	5400
	ATTATTGTTT CAAGTCAGGA AGCAATTAGA GCTGTACCTA ACTCAGTACG CGAACTTCTT	5460
35	ACGGCTTAGG TGCTAATAAA TGGCAAACGA TAAGACGTGT TGTCTTACCA GCAGCGTTAC	5520
	CTGCTATTTT AACTGGATTC ATTTTGTCTC TTTCAAGAGC ACTGGGAGAA ACAGCGCCAC	5580
40	TTGTGCTAAT CGGTATACCG ACTATATTAT TGGCAACACC TAGAAGTATA TTGGATCAAT	5640
	TTTCAGCATT ACCTATCCAA ATATTTACTT GGGCGAAAAT GCCTCAAGAA GAATTCCAGA	5700
	ATGTTGCATC GGCAGGCATT ATCGTTTTAC TAGTTATCTT AATCTTAATG AATGGCGTTG	5760
45	CGATTATTTT ACGTAACAAA TTTAGTAAAA AATTCTAATT TAAACAATCA ATCTCATTTA	5820
	TCTATTAAAA AGGGAGTTTT AAATATGGCG CAAACACTTG CACAACTAA ACAATATCT	5880
	CAAAGTCATA CGTTTGATGT CTCACAAAGT CATCATAAAA CACCAGATGA TACAACTCA	5940
50	CATTCTGTTA TATATTCAAC ACAAATTTTA GACTTATGGT ATGGCGAAAA TCATGCATTA	6000
	CAAAATATTA ATTTAGATAT TTATGAAAAC CAAATTACTG CCATTATAGG TCCATCTGGT	6060

55

	AAAACAGCTG GTAAAATATT ATATCGAGAT CAAGACATTT TTGATCAAAA ATATTCTAAA	6180
	GAACAATTAC GTACAAATGT GGGCATGGTC TTTCAACAAC CTAATCCATT TCCAAAATCA	6240
5	ATATACGATA ATATTACTTA CGGTCCAAAG ATTCACGGTA TTAAAAATAA AAAAGTTCTT	6300
	GATGAAATCG TTGAGAAATC ATTACGTGGC GCTGCAATTT GGGATGAATT AAAGGATAGG	6360
	TtGCACACAA ATGCAATATAG TTTATCCGGT GGGCAACAAC AACGTGTTTG TATCGCGCGT	6420
10	TGTTTAGCAA TTGAACCTGA AGTCATTTTA ATGGATGAAC CGACATCAGC ATTAGATCCA	6480
	ATCTCAACAT TAAGAGTAGA AGAGTTGGTT CAAGAATAA AAGAAAAGTA TACAATTATT	6540
	ATGGTtACAC ATAATATGCA ACAAGCAGCT CGTGTATCAG ATAAAACTGC ATTTTCTTA	6600
15	AATGGTTATG TCAATGAATA TGATGATACT GATAAAATTT TCTCTAACCC ATCAAACAAG	6660
	AAAACAGAAG ATTATATTTT AGGAAGGTTT GGTTGATATA TAATGGCAAT AATTAGACAA	6720
20	CGATATCAGG AGCAACTTGA TGATTTAATA AAAGAATTAC GTCGGTTAGG TGCaAATGTC	6780
	TATGTGAGTA TTGaAAATGG TATAAAAtCA TTAAGTATTG aCGATAGAGG cTTTGCACGA	6840
	CAAACAGTTA AAAACGATAA ACATATCAAT CAATTAAATT ATGATATTAA TGAGCGAGTT	6900
25	ATCATGTTAA TTACAAAGCA ACAGCCCATT GCGAGTGATT TGGGTATGAT GATTTCTTCA	6960
	TTAAAAATCG CCTCCGATTT AGAAAGAATA GGAGATAATG CCTCGAGTAT TGCCAATATT	7020
	CGATTGCGTA CAAAGATTAC AGATGATTAT GTGTAAACCC GTTTAAAGAC AATGGGTAAA	7080
30	TTAGCTATGT TAATGTTAAG GGACTTAGAT CAAGCATTTA AAAAGAAAGA TACCGTATTA	7140
	ATAAGAGAAA TAATTGAGCG TGATGAAGAT ATCGATGACT TATATAGTCA TATTATTAAC	7200
	GCAACGTATC TTATTGATAA CGtCCATTtG TCGCTGCACA AGCTCATTTA GCAGCAAGAC	7260
35	ATTTAGAACG TATTGGTGAT CATATTATTA ACATCGCTGA AAGTGTTTAT TTTTATTTAA	7320
	CAGGtACACA TTACGAACAA TAACTTAAAG TTATTACTAT AAAATCCCTT ACGATAAATA	7380
	TATATTTCTA TTATTCATAA ACCCTCAAAA AAACCAAGAT TCTCACAATT AGTAATGTGA	7440
40	AAATCTTGGT TTATATTGTT CTACTATAAA TTGTCTCGCA TCTTAGTTAT TTGCTTGCTC	7500
	AATTTTCATCT GTTAATTTTT CAAC TTCATC GACTAAATCA GAAATATATT GAATTGTAGA	7560
45	TTTAAGTGGC TGTTCTGTAG TAATGTCTAC ACCTGCAATG TTTGCAAGTT CGACAGGTGA	7620
	TACACTACCA CCTTTTTTCA ATGTTTCTAA CCAAGCATCA ACAGCTGGTT GGCCTTCATT	7680
	TTTAATCTTT TGAGAAACGA CAGTTCCGAT TGTTAAGCCA GCAGAATACG TATACGAATA	7740
50	TAATCCCATA TAGTAATGAG GTTGACGCAT CCATGTTAAT TCAGCACCTT CAGTCATGTC	7800
	TACTGCATCT CCAAAAAATT GTTTATAAAC ATTTAGCATT ATTTCAATTA ATGTnCGGCG	7860

55

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 798 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

```

TTTTTCTTT TCCTCATTG AAAATTGATC ATTCAGCAAT ATAAGCGTAT TTGTTAATGA      60
TTTAGGTGTT CCAATTCAT AATCCCACCA ATTTAAGTTG GTATTCTTGC CAGTTGTTTT      120
AGTAAATTC TCACCTAATT CTTTACTTT TTTATCTGGT TCTTTTCCAT ATGCATTTTT      180
ATGCAGCCAC TCAAGGGCAT CTTTCACTTT CTTCTTATTT TCGTCAGTAT TTAAAGTGGT      240
TTTAGGATTC CTCATCGCTT CTGCGATTTT CTCAATATTA CGATAGGTAC GAGTCATATG      300
AGAAGAATTA GTTCAAGGG TTTCCGCTCC TGACCACAAG TATTTCTTAC CACTTTCAGT      360
TTTCATTTCC TTGAGTAAAT TCGTCGCCTC TTTCTCTGTA GCATCAAAC TCTTCTTCAT      420
ATCTGGATTA TTCTCATCAT ACTTATCATA ACCATAGTTA ACGTCCAGCC ATGTGTTTCT      480
CAATTTTTC TAATCTGGCG TTTGAACATT CGTATCAGCC ACAGCGATTG GATGTTTATC      540
AACACTTCTG AATTCACCAC CATTCAAAGT AATCACACCA GCCATTAATA ACGTAATGGT      600
GGATAATTTT TGCCATTTCT TTATTCTATA TGTCATTGAc ATGTCTCCTT TTTGTGTTGC      660
GCGTGCGCAA TGAATATTAT GATTAAATAA TGATTCAATT TTTCAAAAT CGTTAACGTA      720
TACAAATGAC TGTCTACTGT CAAACAATCC ACAAAGAATG TTGATGtCAT ATaACAATC      780
GATCACCCAA ATTTTCCG                                     798

```

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

```

TACAGGTTTT ACTATAATGG ATGGTATTTT GGCTAAACGA CATTGGTTTA GTCTTCTTTT      60
TTTnACTTCC TAnATTTACA ATGGTATAAA TAATAATGCT ATATTTAGAA TGATGAGTAT      120
ACTTACTGAA ACTAAATTAA AAGTGTCTGG TTCTTTACTA AAGATAGCTG CTATCCTTGC      180

```

	AATACAAGTT CCAATGAGCG CAATTAAAAG TACTAACCCA ACGATGAAAC TCTGTTTGTC	300
	ACTTAACTCA AAGAACTAT AGATAGGATA TTTTTTAATA ATCAAGCCAC CTAAAATCAT	360
5	CCATAAAAAT ACGATAATTC CATAAGTCAC ATTTATAACA TACGTTATTT TTTGGTCACC	420
	AAATCGGACT AATGTATTTT GTAGAATCAG CATACCAATG ACAACACCTA AAATAACGAT	480
	ACTAGCTATA TAAAGTAAAA ATGCAATTGT CACATCAAAT GTACCCAAAT CTAAAAACCT	540
10	AGGAATTAYa AyGACTGCTA AAATAAAAGC GAAGyACAAA GTAATATAkT TATACAAACC	600
	GGTAGTAAGA CTTATCTCAG GTGATAATTG ATCAGCCATT GACTTAATCG GTGTATTAAAT	660
	AATTGAACTT GTATCTTCGT TATTTTTTTC AGCCATAGTT AAATGATCTT CGAGCTCTTC	720
15	CAATAACTCT TCTACTTCTG CTTCACTCTT ACCTCTAAAT AACAATTCAA CACGTAATTT	780
	TTCTAAAAAA TCTTGAGATT GTTTACTTAA CATCGTTTTT CCCTCCAAAC AAGTTAATCA	840
20	TCCCTTTTATT CAAAACCTGC CATTTTCGATT TAAATACTTT TAGTTCCTTT AAACCTGAAT	900
	CGGTAATCGT ATAGTATTTT CGCCTCGGGC CGCCATTACT AGATTTTTTTT ATTGTCGTAT	960
	CAACGTATCC TTTTTTGTTT AAACGCATTA AAACCTGGATA AATACTACCC TCACTTATCT	1020
25	CTGGAAACTC TTGATTCTTA AGTTTCGTCA TAATTTTATA TCCATACGTT TCGCCTTGGG	1080
	CAATGAGACC TAATATCGCC CCATCTAAGA GACCTTTCAT AATCTGATCT GACACTGACA	1140
	TTTTAATCAC CTAATATCTT ACATAATAAG ATAGTACATT GAGAACTTTT CGTCAACTAT	1200
30	CTTTTATTGT AAGGTAGTTG TTGTACACAT TCCTTAAATG ACTAACAAC TGTGTTAATAG	1260
	GGTAATACTT ACGGAAGTAT ATTTTATTTA TGGGGGAGGA ATTAATAATG ACTACAAAAA	1320
	CAGTATTTGA TGTCATTGAT ATGGGGTTAG GATATTTAGT AAATGTGTAT GATGCTTGGA	1380
35	AAGTTGAAAA GGTACTTGAT GATTATCATA AGCCTTTTTT TAATACCATT CATTGGCAAT	1440
	TTGGCATGT ATTAACAATT TTTGAATCGG CCTTAGCTGT TGCTGGTAAA GAGAATATTG	1500
	ATTTAAATAT CTATAGACCT TTATTCGGAA ATGGTTCGTC TCCAGATGAA TGGAAGGATG	1560
40	AAGTACCGAG TATTGAAAGG ATTTTAGAAG GTCTCCAAAC TTTACCTGAA CGTGCACGAA	1620
	ATCTAACTGA AGATGATTTA GCAATTGAAT TGAAACAGCC AATTGTCGGT TGTAATAACT	1680
45	TAGAAGAGTT ATTAGTATTA AATGCCATTC ACATCCCACT TCATGCTGGT AAAATTGAAG	1740
	AGATGTCTCG TATATTAAAA AATTTAAAAT AAATATGTGC TTATTAACCG TTAACAACAC	1800
	GTTAACGGgT TTTTTATTG TTTAAAAGGT CACTTTTTTG AATTTAATAA ACACCATCTA	1860
50	TACCAGTTCT TCACCGATTC TCGAAAAATA ATTATATTAA TGATTTTCGTT AATTTAATTT	1920
	TATATTTAAT TATTACTGTA CATCTTTTGT AGTTAGCTTT ATTCTTAAAT TGAAATATGT	1980

55

	TACTCCCTAT	CGTTGTAGGT	CTCCTTATTT	GGGCACTTAC	ACCTTTTAAA	CCGGATGCTG	2100
	TGGATCCAAC	AGCATGGTAT	ATGTTTCGCA	TATTCGTCGC	GACAATCATT	GCTTGTATTA	2160
5	CACAACCGAT	GCCAATTGGG	GCCGTCTCTA	TAATTGGATT	TACAATCATG	GTACTCGTTG	2220
	GCATTGTTGA	CATGAAAACG	GCTGTCGCTG	GTTTTGGTAA	TAATAGCATT	TGGTTAATTG	2280
	CTATGGCATT	TTTCATTTTCG	AGAGGATTTG	TGAAAACAGG	TCTTGGTAGA	CGTATCGCAC	2340
10	TTCATTTTCGT	CAAATTATTT	GGTAAAAAAA	CATTAGGATT	AGCATATTCT	ATCGTCGGTG	2400
	TAGATTTAAT	TCTAGCGCCT	GCTACACCAA	GTAATACCGC	GCGTGCTGGT	GGAATCATGT	2460
	TCCCAATTAT	CAAATCACTT	TCTGAATCAT	TTGGTTCGAA	ACCGAAAGAC	GGATCAGCAC	2520
15	GCAAAATGGG	TGCATTTCTT	GTTTTCACAG	AATTCCAAGG	TAATTTAATT	ACTGCGGCTA	2580
	TGTTTTTAAC	TGCAATGGCC	GGTAACCCCC	TTGCACAAAA	TTTAGCATCT	AGCACATCTA	2640
20	ATGTTACAT	TACATGGATG	AATTGGTTTC	TAGCTGCTTT	AGTTCCTGGA	CTTGTTTCCT	2700
	TAATTGTTGT	ACCTTTTATT	ATTTATAAAA	TTTATCCACC	AACTGTTAAA	GAAACACCAA	2760
	ATGCTAAGAG	TTGGGCTGAA	AATGAATTAG	CGACTATGGG	TAAAATCGCT	TTAGCTGAAA	2820
25	AATTTATGAT	TGGTATTTTT	GTCGTTGCGT	TAACACTATG	GATTGTCGGA	AGTTTCATTC	2880
	ATATTGATGC	AACTTTAACG	GCCTTTATTG	CGCTAGcATT	gTTATTATTG	ACAGGCGTCT	2940
	TAACATGGCA	AGACATTTTA	AACGAAACAG	GTGCTTGGA	CACATTAGTA	TGGTTCTCAG	3000
30	TATTAGTGTT	AATGGCCGAC	CAATTAAACA	AGCTTGGA	TATTCCTTGG	TTAAGTAAAT	3060
	CCATGCTAC	AAGTCTTGGT	GGCTTAAGCT	GGCCTATAGT	CCTGGTCATT	TTAATATTGT	3120
	TCTACTTCTA	TTACATTAC	TTATTTGCAA	GTTCTACAGC	ACATATCAGT	GCGATGTATG	3180
35	CAGCATTACT	AGgCGTTGCC	ATCGCAGCCG	GTGCACCACC	ATTATTcAGT	GCATTAATGT	3240
	TAGGTTTCTT	CGGTAACCTA	TTAGCTTCAA	CAACACACTA	TAGTAGTGGT	CCAGCGCCGA	3300
40	TTCTATTCTC	TTcAGGTTAC	GTGACTCAAA	AACGTTGGTG	GACAATGAAC	TTAATATTAG	3360
	GTTTCGTCTA	CTTTATTATC	TGGATTGGTT	TAGGATCACT	TTGGATGAAA	GTAATTGGTA	3420
	TATTTTAAAA	TATTTAAATT	AGCGCTCGAA	TCTCAITGAT	TTGGGCGCTT	TTTAATTTGT	3480
45	ATTTAAAATC	AACCTTTGCT	AAATCAAGAC	TCCCTTTTAA	AAATACGTTT	ATCCTTTAAA	3540
	TCATTGCGTG	CTTCACTGAA	AATTTGTATA	AAGATTTAAG	TCATTACGTA	ACATCACATA	3600
	AAATACATTT	CTATACTATT	CCGCTTCATT	GATTAAcATT	ACGTATGCCC	TCATAAATCA	3660
50	TCATACAAAA	AACACCTTCG	TTTAAATTCA	TTTTAATTGC	GAATTCAACG	AAAGTGCCTT	3720
	ATTCATATT	TAATGTTTCA	AATTTATACG	TCTGTCACTG	TTACTGCACA	CATACCTCAG	3780

55

TTATAGGGTT TTTGCGACCG GATGTTTCTT CAATTTAATG TATTGAGAAA GACTATATAA 3900
 CACAATACCT GTCCAAATAA ATATAACGT AATTAATTGA TCTATACTAA AAGGCTCTTT 3960
 5 GAAAACAAAT ATGCCGAGTA CAAACATTAT TGTTGGTCCA ACGTATTGAA TAAATCCTAT 4020
 TAGCGAAAGT GGAATACGTT TTGCCCCGGC TGAGAATAGG ATTAGTGGTA TTGCCGTAAT 4080
 AGCACCAGAA AATAACAACC AAAATGATGA CATGTTCAAT CCAAATGACA TCTGATGTTG 4140
 10 CTGCCATAAA TAAATAACGT ATATTAGTCC AGCAGGTGCG GTAACAATAC ATTCAATCGT 4200
 AATACTGCTG ATGGCATCAA TATGTACTAC TTTTTCAT AATCCGTATG TACCAAAGGA 4260
 TAACGCTAAT ATAATAGAGA CGATTGGGAA TTCTCCAATC TTGAGCGTCA TATATAATAC 4320
 15 ACCGATGAAT GCGAATAAAA TGGCTAGCCA TTCAAATTTA TTGAATCTTT CTTTAAAAA 4380
 GATAAGTGCG AGCAAAATGC TAACAAGTGG ATTTATATAA TAACCTAAAC TTGTTGTAG 4440
 20 GACGTGACCG TTCGTTACAG CCCAAATAAA TGTACCCCAA TTTAATGTAA TGACATAGCC 4500
 TGCTACGACA ATCGCTAATA GCTGAATGGG CTTGCCTAAC AATTGATTCA TATCTCGTTG 4560
 AAATGCATTG CGTTGTTTTT GTCCAACCGC GAGTATGAAA ATCATGAATA TTGCTGAAAA 4620
 25 TATAATACGA AAGGCTAAAA TTTCAAATGC GCCTATTGCA TCAACGAACT GCCAATATAT 4680
 AGGTAGTATT CCCACAGAA TGTATGCACT GAGTGCTAAA AATATGCCTT TTTTATACTC 4740
 TGAATTCACC TTCAAACCTC CTTACTTTCC TAATTTTAA TTTACTGCAT ACGCTCACTT 4800
 30 GGTATGCTA ATATAACGAT TTTACTAATA ATATTTGAT AAAGATATCA TTTTGTTTAT 4860
 ATTTCACACA TTTATTCACC AACCATAA CAATATTAAT TTTATAAATA ATTCTGTACA 4920
 AATCAGGGTA TATTGCCAGA AAGACTACCA TACAACATAA AGGATGGATA CAAATGACTT 4980
 35 TACCTAAAT TGGAAGCCT GCAACACGCG CGCTAAATC ACAAGGTATA TACACATTAG 5040
 AAGCAGTATC ACAATATACG AAGTCATCTC TAATGGAGAT GCATGGCGTT GGTCCATAAG 5100
 40 CTATATCAAT ATTGGAACAA GCTTTATTTC AG 5132

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22243 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

AAGTAAATTA TATTATGAAT TTGCCTGTCA ATTTCTTAAA GACATTCTTA CCGGAACATA 60

	TAGAAGCAAT TAATAATGCT	TAAGAAAAGA CAGCTAATAA TACCGGCTTA AAATTAATAT	180
	TTGCAATTAA TTATGGTGGC	AGAGCAGAAC TTGTTTCATAG TATTAAAAAT ATGTTTGACG	240
5	AGCTTCATCA ACAAGGTTTA	AATAGTGATA TCATAGATGA AACATATATA AACAATCATT	300
	TAATGACAAA AGACTATCCT	GATCCAGAGT TGTTAATTTCG TACTTCAGGA GAACAAAGAA	360
	TAAGTAATTT CTTGATTTGG	CAAGTTTCGT ATAGTGAATT TATCTTTAAT CAAAAATTAT	420
10	GGCCTGACTT TGACGAAGAT	GAATTAATTA AATGTATAAA AATTTATCAG TCACGTCAAA	480
	GACGCTTTGG CGGATTGAGT	GAGGAGTAGT ATAGTATGAA AGTTAGAACG CTGACAGCTA	540
	TTATTGCCTT AATCGTATTC	TTGCCTATCT TGTAAAAAGG CGGCCTTGTG TTAATGATAT	600
15	TTGCTAATAT ATTAGCATTG	ATTGCATTAA AAGAATTGTT GAATATGAAT ATGATTAAAT	660
	TTGTTTCAGT TCCTGGTTTA	ATTAGTGCAG TTGGTCTTAT CATCATTATG TTGCCACAAC	720
20	ATGCAGGGCC ATGGGTACAA	GTAATTCAAT TAAAAAGTTT AATTGCAATG AGCTTTATTG	780
	TATTAAGTTA TACTGTCTTA	TCTAAAAACA GATTTAGTTT TATGGATGCT GCATTTTGCT	840
	TAATGTCTGT GGCTTATGTA	GGCATTGGTT TTATGTTCTT TTATGAAACG AGATCAGAAG	900
25	GATTACATTA CATATTATAT	GCCTTTTTTAA TTGTTTGGCT TACAGATACA GGGGCTTACT	960
	TGTTTGGTAA AATGATGGGT	AAACATAAGC TTTGGCCAGT AATAAGTCCG AATAAAACAA	1020
	TCGAAGGATT CATAGGTGGC	TTGTTCTGTA GTTTGATAGT ACCACTTGCA ATGTTATATT	1080
30	TTGTAGATTT CAATATGAAT	GTATGGATAT TACTTGGAGT GACATTGATT TTAAGTTTAT	1140
	TTGGTCAATT AGGTGATTTA	GTGGAATCAG GATTTAAGCG TCATTTTCGGC GTTAAAGACT	1200
	CAGGTCGAAT ACTACCTGGA	CACGGTGGTA TTTTAGACCG ATTTGACAGC TTTATGTTTG	1260
35	TGTTACCATT ATTAAATATT	TTATTAATAC AATCTTAATG CTGAGAACAA ATCAATAAAC	1320
	GTAAGAGGGA GTTGCTGAGA	TAATTTAATG AATCTCAGAA CTCCTTTTGA AAATTATACG	1380
	CAATATTAAC TTTGAAAATT	ATACGCAATA TTAACTTTGA AAATTAGACG TTATATTTTG	1440
40	TGATTTGTCA GTATCATATT	ATAATGACTT ATGTTACGTA TACAGCAATC ATTTTAAAAA	1500
	TAAAAGAAAT TTATAAACAA	TCGAGGTGTA GCGAGTGAGC TATTTAGTTA CAATAATTGC	1560
45	ATTTATTATT GTTTTTGGTG	TACTAGTAAC TGTTTCATGAA TATGGCCATA TGTTTTTTGC	1620
	GAAAAGAGCA GGCATTATGT	GTCCAGAAAT TGCGATCGGT ATGGGGCCAA AAATTTTTAG	1680
	TTTTAGAAAA AATGAAACAC	TTTACACTAT TAGGTTATTG CCTGTTGGTG GATATGTTCG	1740
50	TATGGCAGGA GATGGCTTAG	AAGAGCCACC AGTCGAGCCC GGTATGAACG TTAATAATTAA	1800
	ACTTAATGAA GAAATGAAA	TAACACATAT CATATTAGAT GATCATCATA AGTTTCAACA	1860

55

	CACTGCTTAT GATAATGAAA GACATCATTT TAAAATTGCT AGAAAGTCTT TCTTTGTTGA	1980
	AAATGGTAGC TTAGTTCAAA TTGCTCCGAG AGACAGACAA TTTGCACATA AAAAGCCATG	2040
5	GCCGAAATTT TTAACATTAT TTGCGGGACC GTTATTTAAC TTTATATTAG CTTTAGTCCT	2100
	ATTTATTGGT CTTGCATATT ATCaAGGcAC GCcTACGTCT ACTGTAGAAC AAGTCGCAGA	2160
	TAAGTATCCA GCTCAACAAG CAGGATTACA AAAAGGTGAT AAGATCGTCC AAATTGGCAA	2220
10	ATATAAAATA TCTGAATTTG ATGATGTTGA TAAGGCGTTA GATAAAGTTA AAGATAATAA	2280
	GACGACTGTT AAATTTGAAC GTGATGGTAA AACAAAGTCA GTTGAATTAA CACCTAAAAA	2340
	GACTGAAAAA AAAGTACTA AAGTAAGTTC AGAGACGAAG TATGTTCTCG GATTCCAACC	2400
15	AGCGAGTGAA CATACACTTT TTAACCAAT TGTATTCCGA TTTAAAAGCT TTTAATCGG	2460
	TAGTACTTAT ATTTTTACAG CTGTAGTAGG TATGTTGGCT AGTATATTTA CGGGCGGATT	2520
20	CTCATTTGAT ATGTTAAATG GTCCGGTTGG TATTTATCAT AACGTCGACT CAGTTGTAA	2580
	AGCGGGTATC ATTAGCTTAA TTGGTtncAC TGC GTTATTA AGTGTAAGT TAGGTATTAT	2640
	GAATTTAATT CCTATTCCTG CACTAGACGG TGGTCGTATT TTATTTGTTA TATATGAAGC	2700
25	GATTTTCAGA AAACCAGTTA ATAAAAAGC GGAAACAACG ATTATTGCTA TTGGTGCCAT	2760
	TTTCATGGTC GTTATAATGA TATTAGTAAC GTGGAATGAT ATTCGACGAT ATTTCTTATA	2820
	ATTTAGGAGG ATAAATAATT ATGAAGCAAT CCAAAGTTTT TATACCAACG ATGCGTGACG	2880
30	TGCCATCAGA AGCAGAAGCA CAAAGTCATC GTTTATTATT GAAATCGGGT TTGATAAAAC	2940
	AAAGTACAAG TGGGATTTAT AGTTATTTAC CGCTAGCAAC ACGTGTGTTA AATAATATTA	3000
	CTGCAATTGT GCGACAAGAA ATGGAACGTA TCGATTCTGT TGAAATTTTA ATGCCAGCGT	3060
35	TACAACAAGC TGAATTATGG GAAGAATCAG GACGTTGGGG TGCATATGGC CCAGAATTAA	3120
	TGC GTTTACA AGATAGaCAT GGAaGACAAT TTgCATTAGG TCCaACACAT GAAGAATTAG	3180
	TTACATCAAT AGTAAGAAAT GAATTGAAAT CATACAAACA ATTACCGATG ACATTATTCC	3240
40	aAATTCAATC TAAATTCCGT GATGAAAAGA GACCACGTTT TGGTTTayTC GTGGGCGTGA	3300
	ATTTATTATG AAAGATGCAT ATTCATTCCA TGCTGACGAG GCATCATTAG ATCAAACGTA	3360
45	TCAAGATATG TATCAAGCGT ATAGCCGTAT TTTTGAGAGA GTTGGCATT ACGCAAGACC	3420
	AGTAGTTGCA GATTCAGGTG CTATAGGCGG TAGCCATaCA CATGAATTTA TGGCATTAAAG	3480
	TGCTATCGGT GAGGATACAA TCGTTTACAG TAAAGAAAGT GATTATGCTG CTAACATCGA	3540
50	AAAAGCAGAA GTCGTTTACG ArcCAaATcA TaAGCATACT ACTGTGCAAC CTTTAGAAAA	3600
	AATTGAAACA CCAATGTTA AGACTGCGCA AGAATTGGCA GACTTCTTAG GTAGACCAGT	3660

55

	GCGTGGCCAT CATGAAATTA ATGACATTAA ATTAAAATCT TATTTGCGCA CAGATAATAT	3780
	TGAATTAGCA ACACAAGACG AAATTGTTAA TTTAGTTGGT GCAAATCCTG GTTCACTAGG	3840
5	TCCTGTAATT GATAAAGAAA TCAAAATTTA TGCAGATAAT TTTGTGCAAG ATTTAAATAA	3900
	TTTAGTTGTC GGTGCTAACG AAGATGGTTA TCACTTAATT AATGTAAATG TAGGTAGAGA	3960
	CTTCAACGTT GATGAATATG GCGATTTCCG TTTTATTTTA GAAGGCGAAA AGTTAAGTGA	4020
10	TGGTTCAGGC GTTGCACATT TTGCTGAAGG TATTGAAGTT GGTCAAGTAT TCAAATTGGG	4080
	TACTAAGTAT TCAGAATCAA TGAATGCTAC ATTCTTAGAT AACCAAGGAA AAGCTCAATC	4140
	TTTAATTATG GGTGTGTTACG GAATTGGAAT TTCTAGAACG CTAAGTGCGA TTGTTGAACA	4200
15	AAATCACGAT GATAATGGAA TTGTTTGGCC TAAATCAGTT ACTCCGTTTG ATTTACATTT	4260
	AATTTCTATT AATCCTAAGA AAGATGATCA ACGAGAACTA GCAGATGCAC TATATGCTGA	4320
20	ATTTAATACT AAATTTGATG TGTTGTACGA TGATCGTCAG GAACGTGCAG GTGTAAATT	4380
	TAATGATGCC GATTTAATTG GTTTACCACT GCGAATTGTT GTTGGTAAAC GTGCATCGGA	4440
	AGGTATTGTA GAAGTTAAAG AACGTTTAAAC AGGTGATAGC GAAGAAGTTC ACATTGATGA	4500
25	CTTAATGACT GTCATTACAA ATAAATATGA TAACCTAAAA TAATTAAGAT CGAATGAATT	4560
	ATAAGAGTAG GAAAAAGCTG AAAGAAATCT GATGCTTATG TCCTGCTCTT ATTATTTTTG	4620
	ATATAATGAT TATTCGATGA AAAATGACTG AAGACATAGT ATAATTAAAG ATAAATTTGT	4680
30	TTTAACAATA TAATGATTAG CCAAATATAA AGCATTTAAT TTTCTATCAT TACTATGCTC	4740
	ACATAATCTA AATATTGTTT GAACACGTAA AAGTAATTTT TATTTAAGGT GGTAATTGTC	4800
	TTGGCAATGA CAGAGCAACA AAAATTTAAA GTGCTTGCTG ATCAAATTAA AATTTCAAAT	4860
35	CAATTAGATG CTGAAATTTT AAATTCAGGT GAACTGACAC GTATAGATGT TTCTAACAAA	4920
	AACAGAACAT GGAATTTTCA TATTACATTA CCACAATTCT TAGCTCATGA AGATTATTTA	4980
	TTATTTATAA ATGCAATAGA GCAAGAGTTT AAAGATATCG CCAACGTTAC ATGTCGTTTT	5040
40	ACGGTAACAA ATGGCACGAA TCAAGATGAA CATGCAATTA AATACTTTGG GCACTGTATT	5100
	GACCAAACAG CTTTATCTCC AAAAGTTAAA GGTCAATTGA AACAGAAAAA GCTTATTATG	5160
45	TCTGGAAGT TATTAAAAGT AATGGTATCA AATGACATTG AACGTAATCA TTTTGATAAG	5220
	GCATGTAATG GAAGTCTTAT CAAAGCGTTT AGAAATTGTG GTTTTGATAT CGATAAAATC	5280
	ATATTCGAAA CAAATGATAA TGATCAAGAA CAAACTTAG CTTCTTTAGA AgCACaTATT	5340
50	CAAGAAGAAG ACGAACAAAG TGCACGATTG GCAACAGAGA AACTTGAAAA AATGAAAGCT	5400
	GAAAAAGCGA AACACAAGA TAACAACGAA AGTGCTGTCTG ATAAGTGTCa AATTGGTAAG	5460

55

	GCAATAGAGG GTGTCAITTT TGATATAAAC TTAAGAAGAAC TTAAGTGG TCGCCATATC	5580
	GTAGAAATTA AAGTGAAGTGA CTATACGGAC TCTTTAGTTT TAAAAATGTT TACTCGTAAA	5640
5	AACAAAGATG ATTTAGAACA TTTTAAAGCG CTAAGTGTG GTAAATGGGT TAGGGCTCAA	5700
	GGTCGTATTG AAGAAGATAC ATTTATTAGA GATTTAGTTA TGATGATGTC TGATATTGAA	5760
	GAGATTAAAA AAGCGACAAA AAAAGATAAG GCTGAAGAAA AGCGTGTAGA ATTCCACTTG	5820
10	CATACTGCAA TGAGCCAAAT GGATGGTATA CCCAATATTG GTGCGTATGT TAAACAGGCA	5880
	GCAGACTGGG GACATCCAGC CATTGCCGTT ACAGACCATA ATGTTGTGCA AGCATTTCCA	5940
	GATGCTCACG CAGCAGCGGA AAAACATGGC ATTAAAAATGA TATACGGTAT GGAAGGTATG	6000
15	TTAGTTGATG ATGGTGTTC GATTGCATAC AAACCACAAG ATGTCGTATT AAAAGATGCT	6060
	ACTTATGTTG TGTTCCGACG TGAGACAACT GGTTTATCAA ATCAGTATGA TAAATCATC	6120
	GAGCTTGCA GCTGTGAAAGT TCATAACGGT GAAATCATCG ATAAGTTTGA AAGGTTTAGT	6180
20	AATCCGCATG AACGATTATC GGAAACGATT ATCAATTTGA CGCATATTAC TGATGATATG	6240
	TTAGTAGATG CCCCTGAGAT TGAAGAAGTA CTTACAGAGT TTAAAGAATG GGTGGCGAT	6300
25	GCGATATTCG TAGCGCATAA TGCTTCGTTT GATATGGGCT TCATCGATAC GGGATATGAA	6360
	CGTCTTGGGT TTGGACCATC AACGAATGGT GTTATCGATA CTTTAGAATT ATCTCGTACG	6420
	ATTAATACTG AATATGGTAA ACATGGTTTG AATTTCTTGG CTAAAAAATA TGGCGTAGAA	6480
30	TTAACGCAAC ATCACCGTGC CATTATGAT ACAGAAGCAA CAGCTTACAT TTTCATAAAA	6540
	ATGGTTCAAC AAATGAAAGA ATTAGGCGTA TTAAATCATA ACGAAATCAA CAAAAACTC	6600
	AGTAATGAAG ATGCATATAA ACGTGCAAGA CCTAGTCATG TCACATTAAT TGTACAAAAC	6660
35	CAACAAGGTC TTAATAATCT ATTTAAATTT GTAAGTGCAT CATTGGTGAA GTATTTCTAC	6720
	CGTAACCTC GAATTCCACG TTCATTGTTA GATGAATATC GTGAGGGATT ATTGGTAGGT	6780
	ACAGCGTGTG ATGAAGGTGA ATTATTTACG GCAGTTATGC AGAAGGACCA GAGTCAAGTT	6840
40	GAAAAAATTG CCAATATTA TGATTTTATT GAAATTCAAC CACCGGCACT TTATCAAGAT	6900
	TTAATTGATA GAGAGCTTAT TAGAGATACT GAAACATTAC ATGAAATTTA TCAACGTTTA	6960
	ATACATGCAG GTGACACAGC GGTATACCT GTTATTGCGA CAGGAAATGC AACTATTTG	7020
45	TTTGAACATG ATGGTATCGC ACGTAAATTT TTAATAGCAT CACAACCCGG CAATCCACTT	7080
	AATCGCTCAA CTTTACCGGA AGCACATTTT AGAACTACAG ATGAAATGTT AAACGAGTTT	7140
50	CATTTTTTAG GTGAAGAAAA AGCGCATGAA ATTGTTGTGA AAAATACAAA CGAATTAGCA	7200
	GATCGAATTG AACGTGTTGT TCCTATTAAA GATGAATTAT ACACACCGCG TATGGAAGGT	7260

55

	CTGCCTCAAA TCGTAATTGA TCGATTAGAA AAAGAATTAA AAAGTATTAT CGGTAATGGA	7380
	TTTGCGGTAA TTTACTTAAT TTCGCAACGT TTAGTTAAAA AATCATTAGA TGATGGATAC	7440
5	TTAGTTGGTT CCCGTGGTTC AGTAGGTTCT AGTTTTGTAG CGACAATGAC TGAGATTACT	7500
	GAAGTAAACC CGTTACCGCC ACACTATATT TGTCCGAACT GTAAAACGAG TGAATTTTTC	7560
	AATGATGGTT CAGTAGGATC AGGATTTGAT TTACCTGATA AGACGTGTGA AACTTGTGGA	7620
10	GCGCCACTTA TTAAAGAAGG ACAAGATATT CCGTTTGAAA CATTTTTAGG ATTTAAGGGA	7680
	GATAAAGTTC CTGATATCGA CTTAAACTTT AGTGGTGAAT ATCAACCGAA TGCCCATAAC	7740
	TACACAAAAG TATTATTTGG TGAGGATAAA GTATTCCGTG CAGGTACAAT TGGTACTGTT	7800
15	GCTGAAAAGA CTGCTTTTGG TTATGTTAAA GGTATTTTGA ATGATCAAGG TATCCACAAA	7860
	AGAGGTGCTG AAATAGATCG ACTCGTTAAA GGATGTACAG GTGTTAAACG TACAACCTGA	7920
20	CAGCATCCAG GGGGTATTAT TGTAGTACCT GATTACATGG ATATTTATGA TTTTACGCCG	7980
	ATACAATATC CTGCCGATGA TCAAAATTCA GCATGGATGA CGACACATTT TGATTTCCAT	8040
	TCTATTCATG ATAATGTATT AAAACTTGAT ATACTTGGAC ACGATGATCC AACAATGATT	8100
25	CGTATGCTTC AAGATTTATC AGGAATTGAT CCAAAAACAA TACCTGTAGA TGATAAAGAA	8160
	GTTATGCAGA TATTTAGTAC ACCTGAAAGT TTGGGTGTTA CTGAAGATGA AATTTTATGT	8220
	AAAACAGGTA CATTTGGGGT ACCAGAATTC GGTACAGGAT TCGTGCGTCA AATGTTAGAA	8280
30	GATACAAAGC CAACAACATT TTCTGAATTA GTTCAAATCT CAGGATTATC TCATGGTACA	8340
	GATGTGTGGT TAGGCAATGC TCAAGAATTA ATTAAAACCG GTATATGTGA TTTATCAAGT	8400
	GTAATTGGTT GTCGTGATGA TATCATGGTT TATTTAATGT ATGCTGGTTT AGAACCATCA	8460
35	ATGGCTTTTA AAATAATGGA GTCAGTACGT AAAGGTAAAG GTTTAACTGA AGAAATGATT	8520
	GAAAEGATGA AAGAAAATGA AGTGCCAGAT TGGTATTTAG ATTCATGTCT TAAAATTAAG	8580
	TACATGTTCC CTAAAGCCCA TGCAGCAGCA TACGTTTTAA TGGCAGTACG TATCGCATAT	8640
40	TTCAAAGTAC ATCATCCACT TTATTACTAT GCATCTTACT TTACAATTCG TGCCTCAGAC	8700
	TTTGATTTAA TCACGATGAT TAAAGATAAA ACAAGCATTG GAAATACTGT AAAAGACATG	8760
	TATTCTCGCT ATATGGATCT AGGTAAAAAA GAAAAAGACG TATTAACAGT CTTGGAAATT	8820
45	ATGAATGAAA TGGCGCATCG AGGTTATCGA ATGCAACCGA TTAGTTTAGA AAAGAGTCAG	8880
	GCGTTCGAAT TTATCATTGA AGGCGATACA CTTATTCCGC CGTTCATATC AGTGCCTGGG	8940
50	CTTGGCGAAA ACGTTGCGAA ACGAATTGTT GAAGCTCGTG ACGATGGCCC ATTTTATCA	9000
	AAAGAAGATT TAAACAAAAA AGCTGGATTA TCTCAGAAAA TTATTGAGTA TTTAGATGAG	9060

55

GAAATAATCA AGGTATTTAT TTAATGCGTA TGGCGTAGTC AAAGAAATAC AAAATTGTTG 9180
 CTGGACACAA AATTATGCCC GTATTTCTTT TCAATGTCTT ACGAGTCTAT TCAAATGTAA 9240
 5 TGGTGAAATA AAGGAACAAA CTTTACAAG AATCTCTGAT TAATAGTGAA GTCATTGTG 9300
 TCAAGCATAA ACTTATGCTA TAATTAAGTT GCTTAAAAAT TAGTGAAGTC AGGCAGAAGA 9360
 GTGGGAGATT CCCGCTCTTT TCTATTTGCC AAAAAGGGAG GCCTGTATGA GTAAAATTAC 9420
 10 AGAACAAGTA GAAGTGATTG TTAAACCAAT TATGGAAGAC TTGAATTTTG AACTTGTTAGA 9480
 CGTTGAATAT GTCAAAGAGG GTAGAGATCA TTTTCTTAGA ATCTCTATTG ATAAAGAAGG 9540
 TGGCGTAGAT TTAAATGATT GTACGCTAGC TTCTGAAAAA ATAAGTGAAG CTATGGATGC 9600
 15 AAATGATCCT ATTCCTGAAA TGTATTATTT AGACGTAGCG TCACCTGGTG CAGAACGTCC 9660
 AATTAAAAAA GAACAAGATT TCCAAAATGC AATAACTAAA CCTGTATTTG TTTCTTTATA 9720
 TGTACCAATT GAAGGTGAAA AGGAATGGTT AGGCATTTTA CAAGAAGTCA ATAATGAAAC 9780
 20 AATTGTAGTA CAAGTTAAAA TCAAAGCAAG AACGAAAGAT ATAGAGATAC CGAGAGACAA 9840
 AATAGCAAAA GCACGTCACG CAGTTATGAT TTAACGTGAT GAGGAGGAAA AAACGTGTCA 9900
 25 AGTAATGAAT TATTATTAGC TACTGAGTAT TTAGAAAAAG AAAAGAAGAT TCCTAGAGCA 9960
 GTATTAATITG ATGCTATTGA AGCAGCTTTA ATTACTGCAT ACAAAAAGAA TTATGATAGT 10020
 GCAAGAAATG TCCGTGTGGA ATTAAATATG GATCAAGGTA CTTTCAAAGT TATCGCTCGT 10080
 30 AAAGATGTTG TTGAAGAAGT ATTTGACGAC AGAGATGAAG TGGATTTAAG TACAGCGCTT 10140
 GTTAAAAACC CTGCATATGA AATTGGTGAT ATATACGAAG AAGATGTAAC ACCTAAAGAT 10200
 TTTGGTCGTG TAGGTGCTCA AGCAGCGAAA CAAGCAGTAA TGCAACGTCT TCGTGATGCT 10260
 35 GAACGTGAAA TTTTATTTGA AGAATTTATA GACAAAGAAG AAGACATACT TACTGGAATT 10320
 ATTGACCCTG TTGACCATCG TTATGTATAT GTGAATTTAG GTCGTATCGA AGCTGTTTTA 10380
 TCTGAAGCAG AAAGAAGTCC TAACGAAAAA TATATTCCTA ACGAACGTAT CAAAGTATAT 10440
 40 GTTAACAAAG TGGAACAAAC GACAAAAGGT CCTCAAATCT ATGTTTCTCG TAGCCATCCA 10500
 GGTTTATTAA AACGTTTATT TGAACAAGAA GTTCCAGAAA TTTACGATGG TACTGTAATT 10560
 GTTAAATCAG TAGCACGTGA AGCTGGCGAT CGCTCTAAAA TTAGTGTCTT CTCTGAAAAA 10620
 45 AATGATATAG ATGCTGTTGG TGCATGTGTT GGTGCTAAAG GCGCACGTGT TGAAGCTGTT 10680
 GTTGAAGAGC TAGGTGGTGA AAAAATCGAC ATCGTTCAAT GGAATGAAGA TCCAAAAGTA 10740
 50 TTTGTAAAAA ATGCTTTAAG CCCTTCTCAA GTTTTAGAAG TTATTGTTGA TGAAACAAAT 10800
 CAATCTACAG TAGTTGTTGT TCCTGATTAT CAATTGTCAT TAGCGATTGG TAAAAGAGGA 10860

55

	GATGCGCGTG	AAGCGGGTAT	CTATCCAGTA	GTTGAAGCTG	AAAAAGTAAC	TGAAGAAGAT	10980
	GTTGCTTTAG	AAGATGCTGA	CACAACAGAA	TCAACCGAAG	AGGTAAATGA	TGTTTCAGTT	11040
5	GAAACAAATG	TAGAGAAAGA	ATCTGAATAA	TAGGTTGGAG	TGAAGTATCT	ATGAAAAAGA	11100
	AAAAAATTCC	GATGCGAAAA	TGTATTCTTT	CAAATGAAAT	GCATCCCAA	AAAGATATGA	11160
	TTCGTGTTGT	TGTTAATAAA	GAAGGCGAAA	TCTTTGCGGA	TGTTACTGGA	AAGAAACAAG	11220
10	GCCGTGGCGC	ATATGTTTCT	AAAGATGTTG	CTATGGTTGA	AAAAGCACAA	CAAAAAGAAA	11280
	TTTTAGAAAA	ATATTTTAAA	GCATCTAAAG	AGCAATTGGA	TCCTGTTTAC	AAAGAAATTA	11340
	TTAGATTAAT	TTATAGAGAA	GAGATCCCAA	AATGAGTATA	GATCAAATAT	TAAACTTTTT	11400
15	AGGATTAGCA	ATGAGAGCTG	GTAAAGTAAA	AACAGGTGAA	TCAGTCATTG	TTAATGAGAT	11460
	TAAAAAGGA	AATTTGAAGC	TCGTTATTGT	TGCAATGAT	GCGTCTGATA	ATACAGCTAA	11520
20	ATTAATTACA	GATAAATGTA	AGAGTTACAA	AGTTCCATTG	AGAAAGTTTG	GAAATCGAAA	11580
	TGAATTGGGA	ATAGCACTTG	GAAAAGGTGA	GCGTGTTAAT	GTAGGGATTA	CTGACCCAGG	11640
	CTTTGCTAAA	AAGTTGCTAT	CAATGATAGA	TGAATATCAT	AAGGAGTGAT	TATATGAGTA	11700
25	AACAAAGAAT	TTACGAATAT	GCGAAAGAAT	TAAATCTAAA	GAGTAAAGAG	ATTATAGATG	11760
	AGTTAAAAAG	CATGAATATT	GAGGTTTCAA	ATCATATGCA	AGCTTTGGAA	GATGACCAAA	11820
	TTAAAGCATT	AGATAAAAAG	TTCAAAAAAG	AACAAAAGAA	CGACAATAAA	CAAAGCACTC	11880
30	AAAATAATCA	CCAAAAATCA	AACAATCAAA	ACCAAAATAA	AGGGCmACAA	AAAGATAACA	11940
	AAAAGAATCm	ACAACAAAAT	AATAAAGGCA	ACAAAGGCAA	TAAAAAGAAT	AATAGAAATa	12000
	ATAAGAAAAA	TAACAAGAAT	AATAAACCAC	AAAATCAACC	AGCTGCTCCA	AAAGAAATAC	12060
35	CATCAAAAGT	GACATATCAA	GAAGGTATTA	CAGTAGGCGA	ATTGCGGAT	AAATTAAATG	12120
	TTGAATCATC	AGAAATTATC	AAAAAATTAT	TCTTACTTGG	TATTGTTGCT	AATATCAATC	12180
	AATCATTAAA	TCAAGAAACA	ATCGAATTAA	TTGCCGATGA	TTATGGCGTT	GAGGTTGAAG	12240
40	AAGAAGTTGT	GATTAATGAA	GAAGACTTAT	CAATCTATTT	CGAAGACGAA	AAAGATGATC	12300
	CAGAGGCAAT	TGAGAGACCA	GCAGTTGTAA	CAATTATGGG	ACATGTTGAC	CATGGTAAAA	12360
	CGACTTTATT	AGATTCAATT	CGTCATACAA	AAGTTACAGC	AGGTGAAGCA	GGCGGAATCA	12420
45	CTCAACATAT	TGGTGCATAT	CAAATTGAAA	ACGATGGCAA	AAAAATCACT	TTCTTAGATA	12480
	CACCGGGACA	TGCTGCATTT	ACAACGATGC	GTGCGCGTGG	TGCaCAAGTA	ACAGATATTA	12540
50	CTATTTTAGT	AGTAGCAGCT	GACGATGGTG	TTATGCCACA	AACAATTGAA	GCAATTAACC	12600
	ATGCTAAAGA	AGCAgAAGTA	CCAATTATTG	TTGCAGTAAA	TAAAATTGAT	AAACCAACTT	12660

55

	GCGGCGAAAC AATTtTCGTc CACTTTCTGC ATTAAGTGGT GATGGTATCG ACGATTATT	12780
	AGAAATGATA GGATTAGTTG CAGAAGTTCA AGAACTTAAA GCAAATCCTA AAAACCGTGC	12840
5	TGTTGGTACA GTTATCGAAG CTGAATTAGA TAAATCACGT GGTCCTTCTG CATCATTATT	12900
	AGTACAAAAC GGTACATTAA ATGTTGGTGA TGCGATTGTA GTTGGTAATA CTTACGGCCG	12960
	TATCCGTGCA ATGGTTAATG ACTTAGGTCA AAGAATCAAA ACGGCTGGTC CATCAACGCC	13020
10	TGTTGAAATT ACAGGTATTA ATGATGTGCC ACAAGCTGGG GATCGCTTTG TTGTATTIAG	13080
	TGATGAAAAA CAAGCTCGTC GTATTGGTGA ATCAAGACAC GAAGCTAGCA TTATACAACA	13140
	ACGTCAAGAA AGTAAAAATG TTTCATTAGA TAACCTGTTT GAACAAATGA AACAAGGTGA	13200
15	AATGAAAGAT TTAAACGTTA TTATTAAAGG TGATGTTCAA GGTTCTGTG AAGCTTTAGC	13260
	TGCATCATTA ATGAAAATTG ATGTTGAAGG CGTAAATGTT CGTATCATTC ATACAGCGGT	13320
20	TGGTGCAATT AATGAGTCAG ACGTGACACT TGCTAATGCC TCAAATGGTA TTATCATTTG	13380
	TTTCAATGTT CGTCCAGACA GTGGTGCAA ACCTGCTGCA GAAGCTGAAA ATGTTGATAT	13440
	GCGTTTACAC AGAGTTATTT ATAATGTTAT CGAAGAAATT GAATCAGCGA TGAAAGGTTT	13500
25	ACTTGATCCA GAATTGAAG AACAAGTTAT CGGACAAGCT GAAGTCGTC AAACATTCAA	13560
	AGTTTCTAAA GTTGGTACTA TTGCTGGATG TTATGTTACT GAAGGTAAAA TTACGCGAAA	13620
	TGCTGGTGTA CGTATTATTC GTGATGGTAT TGTTCAATAT GAAGGCGAAT TAGATACACT	13680
30	TAAACGTTTC AAAGATGATG CTAAGGAAGT TGCAAAGGT TATGAATGTG GTATTACAAY	13740
	TGAAAACTAC AATGACCTTA AAGAAGGCGA TGTTATCGAA GCATTGAAA TGGTTGAAAT	13800
	TAAGCGTTAA TTAAATAAAT TACAAGCTAA AAGTATAGTT AAGATTGATA TGCTCCCTAT	13860
35	AAATATTGCA CTTTTTAAGT GTCTACTTTA TAGGGAGCAT ATTTGATACT AGCTTTTGGT	13920
	TTTTTATTAG AATAGATTAC CTATTAAGG TTACGTTATA TGGACATGAT TTTGTATAAA	13980
	ATTTTGTGGT GGCCTAGAAT GATTTTAAAT GACAAAATAT AATGTCGACT ATTATTGGAA	14040
40	AATTTTCTGT TGaAATGCCT ATCTTACGGC AAACTTTATT TGATTTTATA GGCTTAATTT	14100
	ATTAAAATAA CGTGTGAGCT AAAATAATTG TTTAAGCATT GTTACACTAA AAAATGCAAA	14160
45	TAACAATTGA ACTTAAAGAT AAAGAGGTGA CAAGAATGAG CAGTATGAGA GCAGAGCGTG	14220
	TTGGTGAACA AATGAAGAAG GAATTAATGG ATATCATCAA CAATAAAGTC AAAGATCCTC	14280
	GAGTTGGTTT TATTACAATT ACAGATGTTG TTTTAACAAA TGATTTATCG CAGGCTAAAG	14340
50	TATTTTAAAC TGTATTAGGT AACGATAAAG AAGTAGAAAA TACATTTAAA GCACTTGATA	14400
	AAGCAAAAGG CTTCATTAAG TCTGAATTAG GTTCTAGAAT GCGATTACGT ATTATGCCGG	14460

55

	AAGATTTACA CAAACAAGAT AGATAATTTA GTGTTAGGTA TCTGGAAAAT GTTTGATAAT	14580
	TTCTTAATAT CGGTATATTA ACATTAAACA GTTAATACAT AGATGTGTAG AAATAGTTAA	14640
5	CATTTTCCAG TTTTTTTATG AATAAATTTA GTTGATACGC TATTAAAATA TATTTTAAAA	14700
	AAGAAGGTGA CTATATGTAT AATGGGATAT TACCAGTATA TAAAGAGCGC GGTTTAACAA	14760
	GTCATGACGT TGTATTCAAA TTGCGTAAAA TATTAAAAAC TAAAAAATA GGTCACACGG	14820
10	GTACGCTTGA TCCCGAAGTT GCAGGCGTGT TACCGGTATG TATAGGTAAT GCAACGAGAG	14880
	TTAGTGATTA TGTTATGGAT ATGGGCAAAG CTTATGAAGC AACTGTATCG ATAGGAAGAA	14940
	GTACAACGAC TGAAGATCAA ACGGGTGATA CATTGGAAAC AAAAGGTGTA CACTCAGCAG	15000
15	ATTTTAATAA GGACGATATT GACCGATTGT TAGAAAGTTT TAAAGGTATC ATTGAACAAA	15060
	TTCCGCCGAT GTACTCATCC GTCAAAGTAA ATGGTAAAA ATTATATGAA TATGCGCGTA	15120
20	ATAATGAAAC AGTTGAAAGA CCAAAGCGTA AAGTtAATAT TAAAGACATT GGGCGTATAT	15180
	CTGAATTAGA TTTTAAAGAA AATGAGTGTC ATTTTAAAT ACGCGTCATC TGTGGTAAAG	15240
	GTACATATAT TAGAACGCTA GCAACTGATA TTGGTGTGAA ATTAGGCTTT CCGGCACATA	15300
25	TGTCGAAATT AACACGAATC GAGTCTGGTG GATTTGTGTT GAAAGATAGC CTTACATTAG	15360
	AACAAATAAA AGAACTTCAT GAGCAGGATT CATTGCAAAA TAAATTGTTT CCTTTAGAAT	15420
	ATGGATTAAA GGGTTTGCCA AGCATTAAAA TTAAAGATTC GCACATAAAA AAACGTATTT	15480
30	TAAATGGGCA GAAATTTAAT AAAATGAAT TTGATAACAA AATTAAAGAC CAAATTGTAT	15540
	TTATTGATGA TGATTCAGAA AAAGTATTAG CAATTTATAT GGTACACCCT ACAAAGAAT	15600
	CAGAAATTAA ACCTAAAAAA GTCTTTAATT AAAGGAGATA GAATTTATGA AAGTCATAGA	15660
35	AGtGACACAT CCTATACAAT CTAACAGTA TATTACAGAG GATGTTGCAA TGGCATTTCGG	15720
	ATTTTTCGAT GGCATGCATA AAGGTCATGA CAAAGTCTTT GATATATTAA ACGAAATAGC	15780
	TGAGGCACGC AGTTTAAAAA AAGCGGTGAT GACATTTGAT CCGCATCCGT CTGTCGTGTT	15840
40	GAATCCTAAA AGAAAACGAA CAACGTATTT AACGCCACTT TCAGATAAAA TCGAAAAAAT	15900
	TAGCCAACAT GATATTGATT ATTGTATAGT GGTTAATTTT TCATCTAGGT TTGCTAATGT	15960
	GAGCGTAGAA GATTTTGTTG AAAATTATAT AATTAAAAAT AATGTAAAAG AAGTCATTGC	16020
45	TGGTTTTGAT TTTACTTTTG GTAAATTTGG AAAAGGTAAT ATGACTGTAC TTCAAGAATA	16080
	TGATGCGTTT AATACGACAA TTGTGAGTAA ACAAGAAATT GAAAATGAAA AAATTTCTAC	16140
50	AACTTCTATT CGTCAAGATT TAATCAATGG TGAGTTGCAA AAAGCGAATG ATGCTTTAGG	16200
	CTATATATAT TCTATTAAAG kCACTGTAGT GCAAGGTGAA AAAAGGGGAA GAACTATTGG	16260

55

	TGCTGTTAGT ATTGAAATCG GCACTGAAAA TAAATTATAT CGAGGGGTAG CTAACATAGG	16380
	TGTAAAGCCA ACATTTTCATG ATCCTAACAA AGCAGAAGTT GTCATCGAAG TGAATATCTT	16440
5	TGACTTTGAG GATAATATTT ATGGTGAACG AGTGACCGTG AATTGGCATC ATTTCTTACG	16500
	TCCTGAGATT AAATTTGATG GTATCGACCC ATTAGTTAAA CAAATGAACG ATGATAAATC	16560
10	GCGTGCTAAA TATTTATTAG CAGTTGATTT TGGTGATGAA GTAGCTTATA ATATCTAGAG	16620
	TTGCGTATAG tTATATAAAC AATCTATACC ACACCTTTTT CTTAGTAGGT CGAATCTCCA	16680
	ACGCCTAACT CGGATTAAGG AGTATTCAAA CATTTTAAGG AGGAAATTGA TTATGGCAAT	16740
15	TTCACAAGAA CGTAAAAACG AAATCATTAA AGAATACCGT GTACACGAAA CTGATACTGG	16800
	TTCACCAGAA GTACAAATCG CTGTACTTAC TGCAGAAATC AACGCaGTAA ACGAACACTT	16860
	ACGTACACAC AAAAAAGACC ACCATTACAG TCGTGGATTA TTAAAAATGG TAGGTCGTCTG	16920
20	TAGcATTTaT TAAACTACTT ACGTaGTAAA GATATTCAAC GTTACCGTGA ATTAATTAAA	16980
	TCACTTGSTA TCCGTCGTTA ATCTTAATAT AACGTCTTTG AGGTTGGGGC ATATTTATGT	17040
	TCCAACCTTA ATTTATATTA AAAAAGCTTT TTACAAATAT TAACATTTAT TATATGTTAA	17100
25	GCTAATATTG AGTGAATAAT AAGGTTACAA TGAGATAAAG ATGATATAAG TACACCTAGA	17160
	GTAATAATCA AGATATTAAA AATAAAGTAT GTTTTTTTAA AAAATATAAC TTATATTTAT	17220
	ACTGATAAGG GTGGGACGAT AAGTCTATTT TGTAATAAT AGATGGATAT CCCGCTCTCT	17280
30	TTTTTCCAA TTCAATATTT TATACTAAT ATTAAAATAC GATAATAAAT GATATGATAT	17340
	AACTATTAGA TTCAAGAGAG GAGATTTATA ATGTCTCAAG AAAAGAAAGT TTTTAAACT	17400
	GAATGGGCAG GAAGATCTTT AACGATTGAA ACAGGGCAAT TAGCTAAACA AGCAAATGGC	17460
35	GCTGTATTGG TTCGTTATGG AGATACAGTC GTGTTATCGA CGGCAACTGC ATCAAAAGAA	17520
	CCTCGTGATG GAGATTTCTT CCCATTAACA GTGAACTATG AAGAAAAAAT GTACGCTGCG	17580
40	GGTAAAAATC CTGGTGGATT TAAAAAGAGA GAAGGACGTC CTGGTGACGA TGCAACATTA	17640
	ACTGCGCGAT TAATTGATAG ACCAATTAGA CCTTTATTCC CTAAAGGATA TAAGCATGAT	17700
	GTTCAAATTA TGAACATGGT ATTAAGTGCA GATCCTGATT GTTCACCACA AATGGCTGCA	17760
45	ATGATTGGTT CATCTATGGC GCTTAGTGTG TCGGATATTC CATTCCAAGG GCCAATCGCC	17820
	GGTGTAATG TGGGTTATAT TGACGGTAAA TATATCATTa ACCCAACAGT AGAAGAAAAA	17880
	GAAGTTTCTC GTTTAGACCT TGAAGTAGCT GGTCATAAAG ATGCGGTAAA CATGGTAGAG	17940
50	GCAGGCGCTA GTGAGATTAC TGAACAAGAA ATGTTAGAGG CGATTTTCTT TGGTCATGAA	18000
	GAGATTCAAC GTTTAGTTGA TTTCCAACAA CAAATCGTCG ACCACATTCA ACCTGTTAAA	18060

55

	GAAGAAAAAG	GACTTAAAGA	AACAGTTTTA	ACATTTGATA	AACAACAACG	AGATGaAAAT	18180
	CTTGATAACT	TAAAAGAAGA	AATCGTCAAT	GAATTTATCG	ATGAAGAAGA	TCCAGAGAAT	18240
5	GAaTTACTTA	TTAAAGAAGT	TTATGCAATT	TTAAATGAAT	TAGTGAAAGA	AGAAGTTCGA	18300
	CGTTTAATTG	CAGATGAAAA	AATTAGACCA	GACGGCCGTA	AACCTGATGA	AATCCGTCCA	18360
	TTAGATTCTG	AAGTTGGTAT	TTTACCTAGA	ACGCATGGTT	CAGGTCTATT	TACACGTGGT	18420
10	CAGACTCAAG	CACTTTCAGT	TTTAACATTA	GGTGCTTTAG	GCGATTATCA	ATTAATTGAT	18480
	GGTTTAGGAC	CTGAAGAAGA	AAAAAGATTC	ATGCATCATT	ACAACTTCCC	GAATTTTTCA	18540
15	GTAGGTGAAA	CTGGTCCAGT	ACGTGCGCCA	GGTCGTCGTG	AAATTGGACA	TGGTGCGTTA	18600
	GGTGAAAGAG	CATTAAAATA	TATTATTCCT	GATACTGCTG	ATTTCCCAT	TACAATTGAT	18660
	ATTGTAAGTG	AGGTACTTGA	ATCAAATGGT	TCATCATCTC	AAGCGTCAAT	TTGTGGATCA	18720
20	ACATTAGCAT	TAATGGATGC	GGGCGTACCG	ATTAAAGCAC	CAGTTGCTGG	TATTGCTATG	18780
	GGCCTTGTTA	CACGTGAAGA	TAGCTATACG	ATTTTAACTG	ATATCCAAGG	TATGGAAGAT	18840
	GCATTAGGTG	ATATGGACTT	TAAAGTCGCT	GGTACTAAAG	AAGGTATTAC	AGCAATCCAA	18900
25	ATGGATATTA	AAATTGACGG	TTTAACGCGT	GAAATTATCG	AAGAGGCTCT	AGAACAAGCG	18960
	AGACGTGGTC	GTTTAGAAAT	AATGAATCAT	ATGTTACAAA	CAATTGATCA	ACCACGTACT	19020
30	GAATTAAGTG	cTTACGCGCC	AAAAGTTGTA	ACTATGACAA	TTAAACCAGA	TAAGATTAGA	19080
	GATGTTATCG	GACCTGGTGG	TAAAAAAATT	AACGAAATTA	TTGATGAAAC	AGGTGTTAAA	19140
	TTAGATATTG	AACAAGATGG	TACTATCTTT	ATTGGTGCTG	TTGATCAAGC	TATGATAAAT	19200
35	CGTGCTCGTG	AAATCATTGA	GGAAATTACA	CGTGAAGCGG	AAGTAGGTCA	AACTTATCAA	19260
	GCCACTGTTA	AACGTATTGA	AAAATACGGT	GCGTTTGTAG	GCCTATTCCC	AGGTAAAGAT	19320
	GCGTTGCTTC	ACATTTTACA	AATTTCAAAA	AATAGAATTG	AAAAAGTGGA	AGATGTATTA	19380
40	AAAATCGGTG	ACACAATTGA	AGTTAAGATT	ACTGAAATTG	ATAAACAAGG	TCGAGTAAAT	19440
	GCTTCACATA	GAGCATTAGA	AGAATAATAT	TTAAAGTCAT	ATGACGACAA	TGTATCGTCA	19500
45	TGTGATTTTT	TTATGCCACT	TTTTACGAAG	TGACCCGTTT	TGAATTTGTT	GTATTGAACA	19560
	TTTTAAAACG	CTTTATTATT	TTGTGTGCAA	CTGTTAATTA	TCCTGTATGT	ATAGTGATTA	19620
	ATAGTGATCA	TCAAGTGTTT	TTTAACTTAT	AATGAATAGT	GAGTTTATAT	ATGGACGGGT	19680
50	AACAAATTTA	GGAGGTAAGA	TTTTGAGTTT	AATAAAGAAA	AAGAATAAAG	ATATTTCGCAT	19740
	TATACCATTA	GGCGGTGTTG	GCGAAATTGC	TAAAAATATG	TATATCGTTG	AAGTAGACGA	19800
	TGAAATGTTT	ATGTTAGATG	CTGGACTTAT	GTTTCCAGAA	GACGAAATGC	TAGGTATTGA	19860
55							

	CCTTACACAC	GGACATGAGC	ACGCGATTGG	TGCAGTGAGT	TATGTTTTAG	AACAATTAGA	19980
	TGCACCAGTA	TATGGATCTA	AATTGACAAT	AGCGTTAATT	AAAGAAAATA	TGAAAGCCCG	20040
5	TAATATTGAT	AAAAAAGTTC	GCTACTATAC	AGTTAATAAT	GATTCAATTA	TGAGATTCAA	20100
	AAACGTGAAT	ATTAGTTTTCT	TTAATACGAC	ACACAGTATT	CCTGATAGTT	TAGGTGTTTG	20160
	TATTCACACT	TCATATGGTG	CCATTGTGTA	TACAGGTGAA	TTTAAGTTTG	ACCAAAGTTT	20220
10	ACATGGACAT	TATGCACCAG	ATATTAAACG	TATGGCAGAG	ATTGGTGAAG	AAGGCGTATT	20280
	TGTCTTAATC	AGTGATTCTA	CTGAGGCAGA	GAAACCTGGA	TATAATACTC	CGGAAAATGT	20340
	GATTGAACAT	CATATGTATG	ATGCTTTTGC	AAAAGTGC GA	GGTCGCTTGA	TAGTTTCATG	20400
15	TTATGCTTCG	AACTTTATAC	GTATTTCAGCA	AGTTTTAAAT	ATTGCTAGCA	AGCTAAATCG	20460
	TAAAGTGTC A	TTTTTAGGAA	GATCACTTGA	AAGTTCATTT	AATATTGCTC	GTAAAATGGG	20520
20	GTATTTTCGAC	ATTCCTAAAG	ATTTGCTAAT	TCCTATAACA	GAAGTTGATA	ATTATCCTAA	20580
	AAATGAAGTG	ATAATTATAG	CTACTGGTAT	GCAAGGAGAA	CCTGTAGAAG	CCTTAAGTCA	20640
	AATGGCGCAA	CATAAGCATA	AAATTATGAA	TATCGAAGAA	GGCGATTCTG	TATTTTTAGC	20700
25	AATTACGGCT	TCTGCTAATA	TGGAAGTTAT	CATTGCGAAT	AcATTAAATG	AGCtTgTtAC	20760
	GnCTGGCGCA	CATATTATTC	CAAATAACAA	AAAGATTCAT	GCTTCAAGTC	ATGGTTGCAT	20820
	GGAAGAATTA	AAAATGATGA	TTAATATTAT	GAAACCTGAA	TACTTTATTC	CTGTACAAGG	20880
30	TGAATTTAAA	ATGCAGATAG	CACATGCGAA	GCTAGCAGCT	GAAGCAGGTG	TTGCACCAGA	20940
	AAAGATTTTC	CTTGTGGAAA	AAGGAGATGT	CATTAATTAC	AACGGTAAAG	ATATGATATT	21000
35	AAATGAAAAG	GTAAATTCAG	GAAATATTTT	AATAGATGGC	ATTGGTATTG	GGGATGTAGG	21060
	AAATATCGTG	TTGAGAGACC	GTCTCTTTT	AGCAGAAGAT	GGTATCTTTA	TTGCTGTTGT	21120
	AACGTTAGAT	CCTAAAAATA	GACGTATAGC	TGCGGGACCT	GAAATTCAAT	CTCGTGGGTT	21180
40	TGTATATGTA	CGTGAAAGTG	AAGACTTATT	ACGTGAAGCA	GAAGAGAAAG	TACGTGAAAT	21240
	AGTAGAGGCT	GGTTTACAAG	AAAAACGCAT	AGAATGGTCT	GAAATTAAAC	AAAATATGCG	21300
	TGATCAAATT	AGTAACTAT	TATTCGAAAG	TACAAAACGT	CGTCCTATGA	TTATTCCAGT	21360
45	AATTTCTGAA	ATTTAATCAA	AAAGTCATTA	ACATAAAAGA	GGTCAGAACA	AGTCACTGAA	21420
	ATATAATGGT	TGTCATGGAC	AATTTACTTA	TATTTTATGA	TAGTCAATTG	AAGGGGTAAC	21480
50	GATTAATCTG	TTATCTTAAG	TAAATTGATA	CATAGATGAT	ATTGTTCTAA	CCTCTTTCAT	21540
	CGTCTGTTTG	GACTACATAT	TCTAAACATC	AAATAGGAAA	TTATATATAA	TAACGTCGTT	21600
	TTAACTAAGG	CAACATAAGG	AGGTGCGTCA	ATTGGCACAA	GCAAAAAAGA	AATCGACAGC	21660
55							

GATACGTTAT GTCATAGCTA TTTTAGTAGT TGTATTAAATG GTGTTGGGTG TTTTCCAATT 21780
 AGGAATAATA GGTTCGTCTAA TTGACAGCTT CTTTAATTAT TTATTGGGT ACAGTAGATA 21840
 5 TTTAACATAT ATTTTAGTAC TCTTAGCAAC TGGTTTTATT ACATACTCTA AACGTATTCC 21900
 TAmaACTAGA CGAACGGCTG GTTCGATTGT ATTGCAAAT GCATTGCTAT TTGTATCACA 21960
 GTTAGTTTTT CATTTTAATA GTGGTATCAA AGCTGAAAGA GAACCTGTAC TTTCTTATGT 22020
 10 GTATCAGTCA TACCAACACA GTCATTTCCT AAATTTTGGT GCGGTGTAT TAGGCTTTTA 22080
 TTTATTAGAG TTAAGCGTAC CTTTAATTC ATTATTTGGT GTATGTATTA TTACTATTTT 22140
 15 ATTATTATGC TCAAGTGTTA TTTTATTAAC AAACCATCAA CATCGTGAAG TTGCAAAAGT 22200
 TGCCTGGAA AATATAAAAG CTGTTTGG TTCATTAAAT GAA 22243

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5510 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

TTATTAAATnA TTAATATTTT TATTTTAAa AATAAGCGA GGAGCTATCA ATGGAACAAA 60
 30 TTACTTCTGC ACAAATAAT AGAATTAAC AAGCGAACAA GCTAAAmAG AAACGTGAGA 120
 GGGATAAAAC TGGATTAGCT TTAATTGAAG GTGTGCATTT AATTGAAGAA GCTTATCAAA 180
 35 GTGGAATTGT AATTACACAA TTATTGCAA TTGAACCGGC AAGATTAGAT CAGCAAATTA 240
 WCGCATACGC GCAAGAAGTT TTTgAAATAA ACATGAAAGT TGCTGAATCT TTATCAGGTA 300
 CAGTGACACC ACAAGGGTTT TTCGCAATCA TTGAGAAGCC GCATTATGAT ATTTCTAAAG 360
 40 CACAACAAGT ATTGCTCATC GATCGTGTTT AAGATCCTGG AAATTTAGGC ACATTAATTA 420
 GAACTGCGGA TGCTGCTGGA ATGGATGCTG TAATAATGGA GAAGGTACG ACAGATCCTT 480
 ATCAAGATAA AGTGTGCGA GCGAGTCAAG GTAGTGTTTT CCATTGCCA GTTATGACAC 540
 45 AAGATCTCGA TACGTTTATT ACTCAATTTA ATGGTCTGT TTATGGTACA GCACTTGAAA 600
 ACGCAGTGgC ATACAAAGAA GTTACTTCAA GTGATTCTTT TGCATTACTA TTAGGTAATG 660
 50 AGGGAGAAGG TGTTAATCCT GAATTATTAG CACATACTAC ACAAATTTA ATCATACCTA 720
 TTTATGGTAA AGCTGAAAGT TTAAATGTAG CGATTGCAGG TAGTATTTTA CTTTATCATT 780
 TGAAAGGTTG ACCGTGTTGA AAGTTTTCCG ATATAATTAT AATTAATTGT TTAACAGAAC 840

	ATAAATAATT GTTTTAGGGA GAATAATCGT GACTGCAAGT TATTCCAATT ATTTAAAGTC	960
	TTTTCACCTT TTTGGTTACT TAAAGAGATT TAAGTCGGAA AGACAATCCG TTATCAATAT	1020
5	TAAACAAGTG TATGCTTAGG CATAAATTTG GGTGGTACCA CGGAAATGAC TTTCGTCCCT	1080
	TATTTTTTAA GAGGATGAAA GTCTTTTTTT AGTTAAACAA CAAATATGAT AAATAGAAAA	1140
	TGAATAGTTC GAATAGGGAG GTCAGTGACA TATGTCTGAA CAACAAACAA TGTCAGAGTT	1200
10	AAAACAACAA GCGCTTGTAG ATATTAATGA AGCAAATGAT GAACGTGCAC TGCAAGAAGT	1260
	TAAAGTGAAG TACTTAGGTA AAAAAGGGTC AGTTAGCGGA CTAATGAAAT TGATGAAGGA	1320
15	TTTGCCGAAT GAAGATAAAC CTGCGTTTGG TCAAAAAGTG AATGAATTGC GTCAAACAAT	1380
	TCAAAATGAA TTAGATGAAA GACAACAGAT GTTAGTTAAA GAAAAATTAA ATAAGCCaAT	1440
	TGGcTGAAGA AACAATTGAT GTATCATTAC CAGGTCGTCA TATTGAAATC GGTTCAAAGC	1500
20	ATCCATTAAC ACGTACAATA GAAGAAATTG AAGACTTATT CTTAGGTTTA GGTATGAAA	1560
	TTGTGAATGG ATATGAAGTT GAACAAGATC ATTATACTT CGAAATGCTG AATTTACCTA	1620
	AATCACACCC TGCACGTGAT ATGCAAGATA GTTCTATAT TACGGATGAA ATTTTATTAC	1680
25	GTACGCATAC ATCACCAGTG CAGGCACGTa CGATGGAATC ACGTCATGGT CAAGGTCCAG	1740
	TTAAATTAT TTGCCCTGGT AAAGTGATC GTCGTGACTC TGATGATGCG ACACATAGTC	1800
	ATCAATTTAC ACAAATCGAA GGATTAGTTG TTGATAAAAA CGTTAAATG AGTGATTTGA	1860
30	AAGGTACTTT AGAATTGTTA GCTAAGAAAT TATTTGGTGC TGATCGTGAA ATTCGTTTAC	1920
	GTCCAAGTTA CTTCCCATTC ACTGAACCTT CTGTAGAAGT TGATGTGTCA TGTTTTAAAT	1980
35	GTAAAGGAAA AGGTTGTAAT GTGTGTAAAC ACACAGGATG GATTGAAATT TTAGGTGCTG	2040
	GAATGGTACA TCCTAATGTA TTAGAAATGG CTGGTTTTGA TTCTTCAGAG TACTCTGGAT	2100
	TTGcATTTGG TATGGGACCA GACCGTATTG CAATGTTGAA ATATGGTATA GAAGATATTC	2160
40	GTCATTTCTA TACTAATGAT GTGAGATTTT TAGATCAATT TAAAGCGGTA GAAGATAGAG	2220
	GTGACATGTA ATGTTGATAT CAAATGAATG GTTGAAAGAA TATGTAACAA TCGATGATTC	2280
	TGTAAGTAAT TTGGCAGAAC GTATTACGCG CACAGGTATT GAAGTGGATG ATTTAATTGA	2340
45	CTACACAAAA GATATCAAAA ATTTAGTTGT CGGCTTCGTT AAGTCAAAAG AGAAACATCC	2400
	TGATGCTGAT AAATTAAATG TTTGCCAAGT TGATATCGGA GAAGACGAAC CTGTACAAAT	2460
	CGTTTGTGGT GCACCGAACG TTGaTGCAGG ACAATATGTC ATTGTTGCTA AAGTAGGTGG	2520
50	CAGATTGCCT GGTGGTATTA AAATTAAGCG TGCCAAATTA CGCGGTGAAC GTTCAGAAGG	2580
	TATGATTTGT TCGTTACAAG AAATTGGTAT TTCAAGTAAC TATATACCGA AAAGTTTTGA	2640

55

	ATATTTAGAT GATCAAGTAA TGGAATTTGA TTTAACGCCG AATCGTGCAG ATGCTTTAAG	2760
	TATGATAGGT ACTGCTTATG AAGTTGCAGC ATTATATAAT ACAAAAATGA CTAAGCCAGA	2820
5	GACAACATCA AATGAGCTTG ATTTATCTGC AAATGATGAA CTGACTGTGA CAATTGAAAA	2880
	TGAAGATAAA GTACCATATT ATAGTGCACG TGTGTGTTTAC GACGTGACAA TTGAACCCTC	2940
10	GCCAATTTGG ATGCAAGCAC GCTTAATAAA AGCGGGTATA CGTCCTATTA ATAATGTTGT	3000
	TGACATTTCA AATTATGTGT TATTAGAATA CGGTCAACCA TTGCACATGT TTGATCAAGA	3060
	TGCGATTGGT TCACAACAAA TTGTTGTTTCG TCAAGCTAAT GAAGGCGAAA AAATGACAAC	3120
15	ATTAGATGAT ACAGAACGTG AATTATTTAAC GAGCGATATT GTCATTACTA ATGGACAAAC	3180
	TCCAATTGCA TTAGCTGGTG TTATGGGTGG CGATTTTTCGA GAAGTTAAAG AACAAACATC	3240
	AAATATAGTG ATTGAAGGTG CTATTTTGA TCCAGTTTCA ATTCGTCATA CATCAAGACG	3300
20	TTTAAATTTA CGCAGTGAAT CATCTAGTCG TTTTGAAAAA GGAATAGCTA CTGAATTTGT	3360
	AGATGAAGCA GTCGACCGTG CATGTTATTT ATTACAACT TATGCAAACG GAAAAGTGCT	3420
	AAAAGATAGA GTGTCTTCAG GAGAACTTGG TGCATTTATT ACACCAATCG ACATCACTGC	3480
25	TGATAAAATT AATCGCACTA TTGGATTGA TTTGTCACAA AATGATATTG TTAGTATTTT	3540
	TAATCAACTA GGGTTTGATA CAGAAATAAA TGATGATGTT ATTACAGTGC TAGTACCATC	3600
30	ACGTCGTAAA GATATTACAA TTAAAGAAGA TTTAATTGAA GAAGTTGCAC GTATATATGG	3660
	ATACGACGAT ATTCCATCAA CGTTACCTGT CTTTCGATAAA GTTACTAGTG GTCAGCTAAC	3720
	TGATCGCCAA TATAAACTA GAATGGTTAA AGAAGTGTTA GAAGGTGCTG GATTAGACCa	3780
35	AGCTATTACG TATTCGTTAG TTTCTAAAGA AGATGCTACT GcATTTTCGA TGCAACAGCG	3840
	TCAAACAATT GATTTATTGA TGCCAATGAG TGAAGCGCAT GCGTCATTAC GTCAAAGTTT	3900
	ATTACCACAT TTAATCGAAG CGGCATCATA TAATGTGGCA CGCAAAAATA AAGATGTAAA	3960
40	ATTATTTGAA ATCGGCAATG TCTTCTTTGC TAATGGAGAA GGTGAACTAC CAGATCAAGT	4020
	TGAATATTTA AGTGGTATTT TAACTGGAGA TTATGTAGTC AATCAATGGC AAGGTAAGAA	4080
	AGAAACGGTT GATTTCTATT TAGCAAAAGG TGTCGTGGAT CGAGTATCTG AAAAGTTAAA	4140
45	TCTTGAATTT AGTTATCGCC GTGCTGATAT TGaTGGATTA CATCCAGGTC GTACTGCTGA	4200
	AATCTTATTA GAGAATAAAG TTGTTGGTTT TATTGGTGAA TTACATCCAA TATTAGCAGC	4260
50	TGATAATGAT TTA AACGTA CGTATGTTTT TGAGTTGAAT TTTGATGCAT TAATGGCTGT	4320
	GTCGGTAGGT TACATTAATT ACCAGCCAAT TCCGAGATTC CCAGGCATGT CTCGTGACAT	4380
	TGCATTAGAA GTAGATCAAA ATATTCCAGC AGCTGATTTA TTATCAACGA TTCATGCACA	4440

55

AAAAGGTAAA AAATCAATTG CAATACGTTT AAATTATTTA GACACAGAAG AACATTGAC 4560
 AGATGAGCGC GTTTCAAAAG TACAAGCGGA AATTGAAGCA GCATTAATTG AACAAGGTGC 4620
 5 TGTATTAGA TAATGATTTA AACCCCATGT ATAAGGATAT CTGAAGTAGA TTGATATCCC 4680
 TAACATGGGG TTTTATTTTT GGGTTCACCA ATTTGGTTCC AATGCATTTA AAAAGTCAAA 4740
 GAGGAACAGC GGAATACAGA TGATGcTTCG CACAAC TGCA TAAAAGCCTC TAATGATTAA 4800
 10 AAATCAAAGA GGCTTTAAAA TTTTTTGGGC TTTTTCACGA TTTTAAAAAT GCTTTTTTGA 4860
 AATGGTATCT AAACGTGAAA GACCGTATTT TTTTATAATT TTGGCGGCGA TTACATCGAC 4920
 TTTAGCACCG GCACCTTTAG GAATCGTCAT ATTAATATTT TTTGATATTT GATCCATATA 4980
 15 TGTAACAAAT GCGTATCGAG AAATTATGCT TGCCACTGCA ATGGCTAATG ACTTCGATTC 5040
 TCCTTTTGTT TCAAATTTTG TTTTCTTTGG AAGTGGTATA TCTGATAATG CGTAATGGCT 5100
 20 ATACACTTCG CGTTTTGCGA ACTGATCAAT GACGATATAG TCTAATTGAG ACGAATCAAT 5160
 TTTTTCAGT ACATTTTTGA TGGCTTCATT ATGAAGGGCA GCTTTCATTT TTAATTGAGT 5220
 CCAGCCTTTT GCTTGCTGAA TATTATATTT TTCATTGTGT AGTGTTAATA ATGAATGTGG 5280
 25 TATGAAAGTA ACCAATTGCT CAGCAAGTTC TACAATTTTG GTATCGGTTA ATTTTTTTGA 5340
 ATCATCTACA CCCAAAGTTT TTAAATAGG GACATGCTCT TTGGTAACGA AAGCAGCACA 5400
 CACAGTCAAC GGACCAAAGT AATCGCCACT TCCAGCCTCA TCACTACCAA TACAGTTAAA 5460
 30 TTGrTCATAC ATTAaAGTTg TcCagAAAAG AATTAGCCAT ATTTnCCTTT 5510

(2) INFORMATION FOR SEQ ID NO: 166:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9623 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

GnTTTACTT ATAAATTTTA CGGGGGTAAT ATAATACTtA TTTACCTGTA ATATATGATA 60
 45 ATTCTTCAGC GGCAGCTGCG TTGATAGTTC TATGAGAAAT GATACCTAAT CCTTTAACAT 120
 TGGATTCTGA AATAACGATA GAACCATCAC TGTTAACTTT TTCAACAAAT GCTACATGAC 180
 CGTAATGTTG ATCTGCACCA AATTGTCCAG CCTCAAATAC AACAGCAGCA TGACGTTTTG 240
 50 GTGTATGACT TACTTGATAA TCACGGTATT GAGCTCGATT ATTCCAATTA TGTGCATCAC 300
 CTAAATCACC TGAGATAGAT GTACCAAATT GTTTCATACG GTTATATACG TACCAAGTAC 360

55

	ATGAATCATC ATAATCCTTG ATAGAACGTT CATATTTATC TAAATCTGGC ATGCGTTCAT	480
	CGTCAAACCTG AGTTAATTGA TAGTGTTTAA TAATACTGTT TAATTTCTTA GCATAGTTTG	540
5	GATCTGTAGC ATATGTTTTA GATAAGTGTG ATGTTGCATC TTTATAAGAA TCGGCTTCCG	600
	ATTTCCATGT TGGTTTATAA ATTGTTTCGAT TGCCATCAAT ACCATTTTTA ATAAGGTCAG	660
	AGTAATCTTT TAGTGATTCT TTCGTGCTTG GATATTTTCG GAATCCAGCA TTAATACTAT	720
10	ACAATTGATT ACCATCAGCT TCTAATGTGT TAAAAGGAAC AGAATTCCTT TCAAAAGCAC	780
	CTTTGATACC GAATAAATTA TGGTTTGGTG ACWTAGCTAA AGCACTACGA CCTGAGTCAG	840
15	ATTCTAAGAT TGCTTGGGCA ATCATGACAG ACGCATAAAT ATCGTTATCT TGACCAATGC	900
	GATGTGCATC TTTAGCAATT GATTTGACAA ATTGACGTGT ATCTTTTGAG TCAACAACGT	960
	TAAATTGTCC GCTATCATCA TTGTTAGATA TACTAGGATC TGTTTCGAAT AATGATGTTG	1020
20	CACGTGTATC CTTTTGATTA ACATCGTTAT TGAATGATTG AGCAGGTTTA GATTTATGTT	1080
	TCAATTCATC TTGTGTTGGT AACTGTGGAT TCTTTGTATT AGATTTTCA TTTTGTCTT	1140
	TTTTAGATTG AGATGCATAA TCTTTTGTG TTTTCTTGC ATCTTCAC TG TATTGATCCA	1200
25	AAATAGAGTC TAAAGCCGAA TCTGACATTG ATTGATTATC TTTTCGATGAA GATTTTTGAT	1260
	TTGCTTTATC GTCACCTGCT GGTGACTAT TTGATTGATT AGGTTGTGTT GGCTTTGGCG	1320
	AATTTGGTTG CTTATTAGAT GTACTTGGTT TTGTATTGTT TGATTTAGGT GCTTTTTGAT	1380
30	TGTCTGCTTT ATCTTGTTTA GATGATTGCG TATCAGTGTC ATTTTGTATG CTATTGTCAC	1440
	TGTTTTTATT CGAATCATTT GTTGACTTTT CGCCATTACG AGGTTGTTTCG TAATCAGAAA	1500
35	TATCCGAATT TAAATTGAAT AAGTTTGGGA TTAAAGTTGT TAATGAGTAA TTATCATCGT	1560
	ATTTATTTTT GGTAGCAAT TGGTTTATAT TGGTTTGTGG TAAATTCCTA TAAATAAAAT	1620
	CAATGATATT GTTAGAGTCT GAAGTGCTGT CGTCTATAGT TTTAAATTTT TTGTCGTTAT	1680
40	TGTCTTGGTT ACTTGATTAA TTTTGTCTG CTTTATCAAT ATCTTTACTT GTAGTATCCT	1740
	TAGAAGTTTC ATCGTCATTA GATTTTTTTG AATCATGAGA TGTTGTCTTA GCTGTAGTAT	1800
	CTTTTTGAGG TGTATCAGCA TAAGCGTAG GTGAACTAA AGTAGGTAAT ACGAGCGTAG	1860
45	TTGATAGCAA ATAAATTAAA ATTTTATTTT TAGGCATATT TCGTATTCTC CCTTGAAAAA	1920
	TATAATAATT AAGTGTGATA ATAACTATG ATTTGTTATA ATTTATCGTA TGCTGAAAAT	1980
	AGTTGATAGG TATCAATCGA CTAAATATCT TCCAGTAAAT TGATTATACT AATTCACAAC	2040
50	GCAAAAATAA ATTAATTTAC AAAAAATATA TAAAAAATAT GAATAATTCC TACATAGGAG	2100
	TGTGACAATG AAGAACGCAT TTAAATTATT TAAAATGGAT CTGAAGAAAG TAGCTAAGAC	2160

55

	TAAC TTATGG GCAATGTGGG ATCCATATGG CAACACGGGA CACATCAAGG TCGCAGTCGT	2280
	TAATGAAGAT AAAGGCGACA CAATCAGAGG GAAAAAAGTT AATGTCGGTA ATACGATGGT	2340
5	TAATACACTC AAGAAAAATA AAAGTTTTGA TTGGCAGTTT GTAAGTAGAG AGAAAGCTGA	2400
	TCATGAGATA AAAATGGGTA AATATTTTGC AGGTATTTAC ATCCCATCTA AGTTTACACA	2460
10	TGAAATTACA GGGACACTAC GTAAGCAGCC TCAAAAAGCA GATGTAGAAT TTAAGGTGAA	2520
	TCAGAAGATT AACGCTGTTG CGTCTAAGCT AACAGATACT GGTTCGTCAG TTGTCGTTGA	2580
	AAAAGCGAAT GAACAATTTA ATAAAACAGT AACTCGAGCA TTATTAGAAG AAGCTAACAA	2640
15	AGCAGGTTTA ACTATTGAAG AAAATGTGCC GACAATTAAC AAGATAAAAA ATGCGGTATA	2700
	TTCAGCAGAT AAAGCTTTAC CTAAGATTAA TGACTTTGCG AATAAAATTG TATATTTGAA	2760
	TAACCACCAA GCGGATTTAG ATAAATATGC CAATGATTTT AGAAACTAG GAAATTATAA	2820
20	AGGTGATATT TTAGATGCTC AGAAAAAATT AAACGAaGTC AATGGTGCTA TTCCGCAACT	2880
	TAATGAAAAG GCTAAGTTGA TATTAGCTTT AAATAATTAT ATGCCGAAAA TTGAAAAGC	2940
	GTAAATTTT GCAGCTGATG ACGTGCCAGC GCAGTTCCCT AAAATTAATC AAGGACTTAA	3000
25	CATTGCGAGT CAAGGTATTG ATCAAGCTAA TGGACAGTTA AATGATGCCA AAGGCTTCGT	3060
	CACACAAGTT AGAAGTAGAG TCGGTGATTA TCAAGATGCA ATTCGACGCG CGCAAGATTT	3120
30	AAATCGAAGA AACCAGCAAC AGATTCCTCA AAATAGCGCG GCGAACAACG AAACATCAAA	3180
	TAGTGACCTT GCAGCTGGTA ATGGTGTAGC ATCAACGCCA CCAAGTGAC CAAGTGCGGA	3240
	TACTGCACCA AATAATAATG TTACGCAAAA TACCGCACCA AATAGTAATA ATGCGCCTGT	3300
35	ATCGACTACA CCACAAAGTA CAAGCGGGAA AAAAGATGGT CAAAGTTTGT TAGATATAAC	3360
	AACAACACAA GTCAGCACAG CTAACGAGAA CACACAAAAC ATTACAGATA AAGATGTTAA	3420
	ATCAATGGAA GCGGCATTAA CGGGCTCTTT ATTATCATTA TCAAATAATT TAGATACCCA	3480
40	AGCGAAAGCC GCACAAAAAG ATAGTCAGGC ATTACGTAAT ATTTTCGTATG GGATTTTAGC	3540
	ATCGGACAAG CCTTCTGATT TTAGAGAGTC TTTAGATAAT GTTAAGTCCG GTTTAGAATA	3600
	CACAACGCAA TATAATCAAC AATTTATCGA TACATTAAAA GAGATTGAGA AGAATGAAAA	3660
45	TGTTGATTTA TCAAAAGAAA TTGATAAGGT AAAAGCAGCT AATAATCGAA TTAATGAATC	3720
	ATTAAGGTTA GTTAATCAAT TAAGCAATGC ATTAAAGAAT GGTAGTTCAG GAACTGCTGA	3780
50	AGCTACTAAA TTAGTAGATC AACTTTCAAA ACTAGATTCA TCATTATCAT CATTAGAGA	3840
	TTATGTTAAA AAAGATCTTA ACAGCTCTTT AGTATCAATA TCACAACGTA TTATGGATGA	3900
	ATTGAACAAA GGGCAAAC TGTTTCAGTCT AAATTAAATA CAATTGATCA	3960

55

	AACAGTATTA CCAAGTATTG AACAAACAATA CATTAGTGCT GTTAAAAATG CTCAAGCAAA	4080
	CTTCTCGAAA GTGAAAAGTG ATGTAGcTAA AGCTGCTAAC TTTGTGCGCA ATGACTTACC	4140
5	ACAGTTAGAA CAGCGATTAA CTAATGCGAC AGCAAGTGTG AATAAAAAATT TACCAACGTT	4200
	ATTAAATGGT TATGATCAAG CGGTAGGATT ACTAAATAAA AATCAGCCAC AAGCGAAAAA	4260
10	GGCTTTATCA GATTTAGCTG ATTTTCTCA AAATAAATTG CCTGATGTTG AAAAAGATTT	4320
	GAAAAAGCG AATAAAATTT TCAAGAAATT AGACAAAGAT GATGCAGTCG ACAAATTAAT	4380
	CGACACACTT AAGAATGATT TGAAAAAGCA AGCGGGTATT ATTGCAAATC CTATTAATAA	4440
15	GAAGACTGTT GATGTTTTCC CAGTTAAGGA TTATGGTTCA GGTATGACAC CATTCTATAC	4500
	TGCACTGTCA GTATGGGTAG GTGCACTCTT GATGGTAAGT TTATTAACGG TTGATAATAA	4560
	ACATAAGAGT CTAGAGTCAG TCTTAACGAC AAGACAAGTG TTCTTAGGTA AGGCAGGATT	4620
20	CTTTATAATG CTTGGTATGT TGCAAGCACT CATTGTATCG GTTGGAGATT TGTTAATCCT	4680
	AAAAGCAGGA GTTGAGTCAC CTGTATTATT TGTACTTATA ACGATTTTCT GTTCGATTAT	4740
	TTTCAACTCA ATCGTATATA CGTGCGTATC ATTACTTGGT AACCCAGGTA AAGCCATTGC	4800
25	AATCGTATTG CTTGTATTAC AAATTGCAGG TGGTGGGGGA ACATTCCCAA TTCAAACTAC	4860
	GCCACAATTT TTCCAAAACA TTTCGCCATA CTTACCATTT ACGTATGCAA TTGATTCATT	4920
30	ACGTGAAACA GTAGGCGGTA TTGTTCCGGA AATCCTAATT ACAAATTAAT TTATATTAAC	4980
	GTTATTTGGT ATAGGATTCT TCGTTGTAGG TTTAATTTTA AAACCTGTAA CAGATCCATT	5040
	GATGAAGCGC GTATCTGAAA AAGTTGACCA AAGTAACGTT ACAGAATAAA AATTAAATCC	5100
35	ACACATTAGG GTTATAGCTC CTTAATGTGT GGATTTTAT GTTTTATAGAC AGAAGAGATA	5160
	GTAATTTCTG TCTTTTATGG GACGGTTGTT ATCATTGCTA TTATCCAGGA TGAATTAATA	5220
	TAGGACTAAT ATTACCGACA AAGTGAATAT CCTCGTCTTC CGTAGTTAAA ATAAAGCTAG	5280
40	AACCTTTTTG GATGTCATAG TGCTTATCGT TTAATGTTAA AGTACCAGTA CCATCGATAA	5340
	TTGTAACTAA GCAATAAGCA TGTGGTTTAT TGAATTTTAA ATCTCCATGA ATATCCCAT	5400
45	TATATACTGC AAAATATTGA TTATCTACAA ATTGAGTTAC AGTGTGTGTG TCGATGTGAG	5460
	TTGTTATAGG AGTAGTATTT GGTTCATGAT TGCCTAATTC AATCACATCT TTACTTTGCT	5520
	CTAAGTGCAA ATCACGCAAT TGACCATTTT GATCTCGTCT ATCATAGTCA TAAATACGGT	5580
50	ATGTCGTATC GGAGGATTGT TGTGTCTCTA AAATTAAAT ACCCGAACCA ATGGCATGGA	5640
	CAGTGCCAGC AGGAACATAA TAAAAGTCAC CGGGCTTAAC AGGTATACGT TTGAAAAGAC	5700
	TGTCAAATTC ATGATTATCA ATCATGTCTA TTAACGTCTG TTTATTATGT GCATGTACGC	5760

55

	GTTCGCCTTC GTGTTTTAAA GCGTAGTCAT CATCTGGGTG AACTTGAACA GATAATTTAT	5880
	CATTGGCATC TAATACTTTA GTTAGCAGAG GGAACTATC TCGTGAATCA TTATCGAATA	5940
5	ATTACAGATG TTGTGACCAA AGTTGATCTA GGGTCATATC CTTGTATGGA CCATTGATAA	6000
	TTGTATTAGG ACCATTTGGA TGTGCAGAAA TTGCCAGCA TTCACCAGTT GTTTCATTAG	6060
	GGATATCATA GTTAAATGCT TTTAATGCAT GACCGCCCCA AATTCTGTCT TTAAAAACGG	6120
10	GTTGTAAAAA TAATGCCATA GTTAAAACTC CTCTATATTT TCATTAAATA GTTATAAATT	6180
	TCTGTAGTAC TGTTTGCAAT AATTAGTGAT TGGCGTGTCT CATCATTCAT TAACGCTTTA	6240
	GATAAGCGCT GAAGTATTTT TAAATGTGTA TCCTGACTGT TGTTTGGTAC GGCAATTAAG	6300
15	AATATCAATT GAGGTAGACT ACCATCTAGA CTGTCCCATT TAACACCATG ATTATTTTTC	6360
	ATAACAGCTA CAATCGGTTG TTTTACAACA TCAGACTTTG CATGTGGAAT GGCCACGTTT	6420
20	ATGCCAATAG CTGTCTAGm tCcATTTCAC GTTCTAGTAT TGCATTTTTT AAATGCGATG	6480
	TGTGCTCTAC ATAACGGCAA ATTTTAAGTT TATGAATCAA CATATCAATT GCTTCGTTTC	6540
	GAGACATGTC GTGATCAGTA ATTATCATAG TTTGTTGATC AAAACATGA GAAGGTTTAT	6600
25	TGAGATGTGA ATGTTTCGCG GTGTTATCTA CATTGTCAAC CTCTGTATCA TGTGTGTAA	6660
	TATCTGTATC ATGAAGTTGC GTGTGTTGCG CTGGTGCATC TACTGCTATA ACTGGTGTAT	6720
	TGCGTTTTTAA TAATAGTACA GTAGTCATTG TGACAAGACT ACCTACTATC ACTGCAAAGA	6780
30	TAAACCATAA TACATGATCA ATACCACCTA ATACAGCCAC GATTGGACCT CCATGTGCGA	6840
	CTCTATCGCC GACACCACCA ATGGCTGCAA TGAATGATGC AATCATTGCA CCAATGATGT	6900
	TTGCAGGTAT AATGCGCAAT GGATCTTGGG CTGCGAAAGG AATAGCACCT TCAGTAATAC	6960
35	CAAATAGTCC CATAGTGAAG GAAGCCTTAC CCATTTCTCT TTCGGAATGA TTGAATTTAT	7020
	ACTTTTGAAC AAACGTTGCT AAACCTAAAC CGATTGGTGG TGTACATACA GCAACTGCGA	7080
40	CCATACCCAT AACGGCGTAA TTACCTTCAG CAATAAGTGC TGAGCCAAAT AAAAATGCTA	7140
	CCTTGTTTAC TGGACCGCCC ATATCGAAGG CAATCATCGC ACCTATAATC ATCGCAAGTA	7200
	TAATAATATT AGCACCTTGC ATACTTTTTA ACCAGGTTGT TAATGCCTCA AAAATATTAG	7260
45	AAATTGGTGC ACCGATTAAA AATATAAATA TCAATCTAC AACGACCGAT GAAATAATGG	7320
	GAATAATAAT GATAGGCATA ATTGGTGCCA TTGCTTTTGG AACTTTAATA TCTTTAATCC	7380
	ACTTTGCGAT ATAACCTGCT AAGAAACCAG CAACAATACC ACCTAAAAAT CCTGCGCCTG	7440
50	CATCACTGCC ATAAAACTA CCGTCAGCAG CGATAGCGCC GCCAATCATA CCAGGAACAA	7500
	GACCGGGcTT GTCAGCGATA CTAACAGCGA TATATCCAGC TAGTATTGGA ACCATAAATT	7560

55

	ATCCTTTTGA TGTCGTTTCA CCGCCTAGAG TCAGCGCGAT GGCGATAAGG AGTCCACCAA	7680
	CTACGATAAA AGGAACCATA AACGATACAC CGTTCATTAA ATGTTGATAC ACCATTTGAA	7740
5	TACCAITTTT AGACTTACCG CGATCTTTCG AATGATAATT TGTTTCAGAT TGATAAATAG	7800
	GCGCATCTTG ATTAATGATA CGTTGAATTA GACCTCTCGG ATTATGAATC CCTTCGCGAA	7860
10	CATTTTCATT AATCAACCGT TTACCAACAA ATCGGGACAG ATCAACTTGT TTATCAGCTG	7920
	CAATTATGAC ACCGTCAGCT TCTTCGATGT CTTGCGTAGT TAAAACATTT TCAGCACCAA	7980
	CACCGCCCTG TGTCTCTACT TTAATATCCA CACCCATTTT TTTTGCTACC TGCTCAAGCT	8040
15	TTTCTTGAGC CATATATGTA TGTGCAATGC CATTTGGGCA TGAGGTAATA GCTACAATTT	8100
	TCATAAAATC ATCTCCTTTT CTATATTGTA AGCGTATTCT CGATACTAAA AAAAAGAATA	8160
	ATTACCGTTA CTAGTGGCAA TTATTCTTGT AAGTATTCAA ATAAGTGTG CTTTAAACTA	8220
20	TGATCATCTA AACTACATAA ATGGTTCACT GAATCATCAT CCAAGTTAGC AATTAATTGC	8280
	ATCATTGTGT TTGTAAAAGC TTTGTCTTTA TGCGAAATCG CTAAGAAAAA GACAAGTTTG	8340
	ACATCGTGTT GTCGCCAAGG AAAAACATCT TTTGTGCGAA AAATAAGCAC ATGTGATTGT	8400
25	AAAACTTTTT CAGGATCTCC ATGAGGAATC GCCATAAAAT TACCTATGTA TGTAGAAGAT	8460
	GATTTCTCAC GCTCTAAAGC TGATTCGATA TATCCTTCTA CAATCGCATG ATGTGCTTGT	8520
	AATATTTTTT GAGCTTCTTC AAAAATTTGC ACAGTATGCC GTGATTTTTG TTCAGTATTT	8580
30	ACGACAAGGA AATTGACAGT GTCCATATGA TGATGTGCTT GAACCGGATT TTGCTTTTGC	8640
	TTCACAACGT GTCTGATTTT GTGACGATCA TCTTCAGAAA ATAATGGTGC AACCTTGATA	8700
	GTCGTCAGGT GCTTAGGAAG TATGTTTAGC GTTTGTTTAG GAATATCATG GGTCGTTATT	8760
35	AATAAATCTA CATTGTCAA GTGATAGTGT GTTATATTTT CTAGTTTAAT CGTATTTATC	8820
	ACTGACAAC CTTCGGATAA GTTATTTATT TTAGTTTCTA AAAAATTCGA CACACCTAGA	8880
40	CCATAATAAC AAGCAATGAC TACATTTAAT TGTGTTTTGG TACGACGCTC GATGGCAGCT	8940
	TGAAAATGAA TTGTTAAAA TGCAATTTCA TCTTCGCTCA TCTCTATATC AGTATCAATT	9000
	GCTAATTTAT CAATCGCTTC AAAAAGTGTG TTAAACACAA AGGGATAGAG TTTTTTAATC	9060
45	TCTATAACTA AAGGATTGTT TAAATAAATG TTTTGAGTGA TACGTAAATA TGCTTTACTA	9120
	AAATGATTAT ATAAATTTTG TTGTAAAATC GAATCTTCAT TGAAAGGTAC ATGAATACGT	9180
	TGCTGCATCA ATTCGATTAA GCGATCAATA TAACTTTGTA TAAATATACG TTCTATGCCA	9240
50	ATATCGAGTT TATTAAAATG ATAAGCAATA AAGAATGAAA ACATATTGAT TACTTTTTCG	9300
	TTCAAGTCAT AACCTAATCT TTCGTTGATT TGCTTAATGC AAGATTGAGA TATCAATTTT	9360
55		

AGATGAATTA AAAGCTGTTG TATTTGAATA TCAGTTGTTT CAATACTATG TTGTTGAAGT 9480
 GTCTCTTGTA TAATATGCGA AATCATCCTT TGGTGTGAAT CAGGTAATTC aTTTAAAATT 9540
 5 AGGTCTTCAA CATGTACATG CCCTGATGAT AATTGATTTA AATGGATGAT GGCATTAGTG 9600
 ATATCATTAT CTGTTCCATC GAC 9623

(2) INFORMATION FOR SEQ ID NO: 167:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

20 ACCGTGGAAA CACGTCTAGT CAATCAGAAA GCGATAAAAA TGTGACTAAA TCATCTCAAG 60
 AGGAAAATCA AGCAAAAGAA GAATTACAAA GCGTTTTAAA CAAAATTAAC AAACAATCAA 120
 GTAAGAATAA TTAAAAAATT TTGATATTGT CTATGTTTAT AGTTCACAAG CCATTCAACG 180
 25 TATTGTAAAC TAAGGATAGT GTATTTTTTT AATAGTAATT TGTCAGGAGG TGCCTATCTA 240
 TGGAGAACA TTA CTACGTA AGTATTGATA TTGGATCATC AAGCGTAAAA ACAATAGTAG 300
 GCGAGAAATT TCACAATGGT ATAAATGTGA TAGGTACAGG ACAAACCTAC ACGAGCGGTA 360
 30 TAAAAAATGG TTTAATTGAT GATTTTGATA TTGCGCGACA AGCAATCAAA GACACAATTA 420
 AAAAGGCATC AATCGCTTCG GGTGTTGATA TTAAAGAAGT TTCTCTGAAA TTACCTATCA 480
 35 TTGGAACGGA AGTTTATGAT GAATCAAATG AAATCGACTT TTATGAGGAT ACAGAAATCA 540
 ACGGTTTACA TATCGAAAAA GTATTAGAAG GTATTAGAGA AAAAAATGAT GTGCAAGAAA 600
 CAGAAAGTAAT TAATGTGTTT CCGATTCTGT TTATAGTCGA TAAAGAAAAT GAGGTTTCAG 660
 40 ACCCTAAAGA ATTAATTGCC AGACATTCAT TAAAGGTTGA AGCAGGCGTA ATTGCTATTTC 720
 AAAAATCGAT TTTAATTAAT ATGATTAAAT GCGTAGAAGC ATGTGGTGTT GATGTATTAG 780
 ATGTTTACTC TGATGCATAT AACTATGGTT CAATCCTAAC AGCTACTGAA AAAGAGTTAG 840
 45 GTGCATGTGT CATTGATATT GGTGAAGACG TTACGCAAGT TGCTTTTTAT GAACGCGGTG 900
 AATTAGTAGA TGCTGATTCT ATCGAAATGG CAGGGCGTGA TATTACaGAC GATaTTGCAC 960
 aAGGrTTaAA CACTTCTnAT GAAACTGCTG nAAAAAGTTA AACACCAATn TGGTCATGCA 1020
 50 T 1021

(2) INFORMATION FOR SEQ ID NO: 168:

(A) LENGTH: 7963 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

10	TAATCTATTA TAAAACTGT CCATACCCTT TGATTACCTT CTCTTCAGGT ACAGGCCACA	60
	CTTGAGGCCA TAAGCCATAT GCTTGCTGTG AATAAAATTG TGCCATTTGT AACAAATATAA	120
	TATATACAAA TAAACACCCA ATAATTGCTG TCACTAATGG ATATGATAAC CAAACCATTA	180
15	ATAAACTGC AATAATTACT AACCTAAAGA TAATATTAAA TGCGTCTCTC CCTCTTATAA	240
	AGCTTCTAAT AAATAAGAAT AAATACATCG CATTAGAGTT AAATTTACTA CCCTTTGGAA	300
	CTGGTAAAG TATATCTAGA TAACTTCTTC TGACTGCAGA TTCTTTCAAA TGTTTTACAT	360
20	CGGTGAACAT ATTAACAAAT TTATAATAAT TCATATGATG TCGATGTTTCG ATTGCAATCA	420
	TTTTCTCCCA AGGATACAAA AAGCCTGGTT TATATTTTTT AACTAAAAAT TCTATTAACA	480
	CAGGCAAAGC AACCATCACA AATGCGATGT ACCATTTTGG AGCTAATAGT AAGTAATATG	540
25	TTAGAGCAAA GGTGATGAAT GATATTAAAT TAACTTGCCA TGTTTTAAGT CCCGATTGAT	600
	ACCATTGCCA TCTTAAGCGT AAACCAACAT ATGGAAAAAT TAATGCACTG ACTCCAAAAC	660
30	AAATATAAAA TGCCACATTA TGTTGATTAA TATTGTAAAA CAACGGGAAC ATTACAATAA	720
	CAATAATGAG TTGGATTAAT ATGCGCGCAA AGTAACTATA TAAATCGCA TGACGCATAA	780
	ATTGAGACAT GTGTTTTTCA AATGGTAATA AAAAGATTTT ATCCgCTTCT TTTAACAGTG	840
35	GTCsCmTTGG AAAAATAGrT GTCAACGCAA CAATCACTGC TGCTATTaAT GAAAAATTGa	900
	TATTCGTTGG AATATGTTTT AACCATTcAC CATATCCArA AATAAATGCA CCCAGCAAAA	960
	TAAGTAAAAA GACCATGAAA TGACCATTAA ATATAAACTT ATTATAATAA TTTTtCTCTT	1020
40	TACGAAGGGC ATGTAATCTT TTATTAAATA ATGTGGTAgC TTGGTTACGC ATGTACATCT	1080
	CCACCTTGCG TCACATGAAT ATATATATCG TCTAATGTTT GATTATGTAA GCCAGTTTGT	1140
	TGTCTCAATG CTTCTAAATC TCCAAATGCA ACGACTTCAC CTTCGTCTAG TATGaTAAAA	1200
45	CGATCACAGT AACGTTcAGC TGTTGCTAAA ATATGTGTAC TCATTAGAAC GGTTCTACCT	1260
	TCGTTTTTCT TTTCAACCAT TAAATCTAAC ATGGATTGAA TTCCTAATGG ATCTAGGCCA	1320
50	AGGAATGGTT CGTCTATAAT ATACAATTCG GGATTAAACGA TAAACGCACA AATAATCATG	1380
	ACTTTTTGTT TCATCCCCTT AGAAAAATGA CTCGGAAAAA CTTTCAACTC ATTTTCTAAA	1440
	CGGAATGTCT TTAATAATGG CATTGCTCGA TTCATCGTTT CATCACGATC AATATCATAT	1500

55

	TCCGGAATAT AAGATAACTT TCTTCTATAA GCCTCTATGT CATCATTAAT GTTGATATCT	1620
	GAAATTGATA GAGATCCTTC CATAGGTGTA AGCAATCCTA GCATATGTTT AATCGTTGTA	1680
5	CTCTTACCAG CGCCATTAAAG GCCAATAAGT CCAACAATTT CGCCTTTGTT TAATTCAAAA	1740
	TTTATATCTT TAATTACAGG GCGTTTTCCTA TATCCACCTG TAAGCTGTTC TACTTTAACT	1800
	GTCATAAGGC ACCTCCATGA CTTATATTGT ACCAAAAATT ATAAAATGCT CATATTAAAT	1860
10	ACACATGTCC TAATATCGAA TTTTGTAGCGA CAATGTTATA ATGAATGGTA ATACTAGTTG	1920
	AAAAGGAGTG TAGTCATCAT GTCAGAAACA ATTTTCGGCA AAATTTTAAAC TGGAGAAATT	1980
	CCTAGCTTTA AAGTATATGA AGACGATTAT GTCTATGCCT TTTTAGATAT ATCACAAGTT	2040
15	ACTAAAGGAC ATACGTTATT AATTCCTAAA AAAGCTTCTG CTAATATCTT TGAAACTGAT	2100
	GAAGAAACAA TGAAACATAT CGGTGCAGCA TTACCTAAAG TAGCAAATGC TATTAAGCGT	2160
20	GCATTTAATC CTGATGGTTT AAACATTATT CAAAATAATG GTGAGTTTGC AGATCAATCT	2220
	GTATTTTATA TTCAATTTCCA CTTAATTCCT CGATACGAAA ATGATATTGA TGGATTTGGT	2280
	TATAAGTGGG AAACACATGA AGACATTTTA GATAACGATG CAAAACAACA AATTGCTGAA	2340
25	CAAATTCAG CACAATTTTA AATGTATGCT TAATCTAAGC TCGAACGGGT ATAATATGAT	2400
	TAATATTATA ACAATTGCGT TTGAAGTGAT AACATCAAGG TTAGCAATTT TAAACAAAAT	2460
	GAGTTATCAA GATAACAGAT GTTAAAAGTG AGGAGAATAT AAATGAAAGC ATCAGCGATT	2520
30	CTATTCGGTA TCGGTGTTGG CGTAGCAGCT GGTTTTGTAG TTGCACTTCA AGGACGTGAC	2580
	GACAAAAGTG TCAAGAACAA CACGATCGAT CGTACTGCCC CTACTGGTTC AAAATCAGAA	2640
	CTACAACGTG AATTTGAAAC GATTAAACAA AGTTTTAATG ACATTTTAAA CTATGGTGTT	2700
35	CAAATTAATA ACGAAAGTGC GGAATTTGGT AGTTCAATTG GTGGTGAAAT TAAGTCATTA	2760
	CTTGGAAACT TCAAATCTGA CATTAACTCT AATATTGAAC GTTTACAGTC ACACATCGAA	2820
40	AATTTACAAA ATCGTGGCGA GGATATTGGA AACGAAATTT CTAAGTAGCA GGTACGTTT	2880
	TCGATCACAA CTATTTTTAT TAGTAACAGC ATATTTATTT TTTAAAATTA AATGCCAAAT	2940
	AAACGAGATG ACATTAGAAA TTAGATATTT CTTGTCATCT CTTTTTTAAA ACTCAAATGA	3000
45	ACTTATGTTT ACAAATTATA GGAAGACATT GTTTGTAGTG ATTTTCGCTT AAATCATATT	3060
	TATGAATTGA TTGAAAACAT TGCTTAGGAT TCATTGTGTT ATCCTTGCAC TTTGATTACG	3120
	CTTTACTTAA ATCATTATCG ACAAACAACA TACTTATATT TTCATTGAGC CGAACCTTAT	3180
50	ATACACATTA CATATACCTT ACTTGACAA ATTATTAATC TGGTGTATTA TATAATTACA	3240
	TATCACTATA TTTTGTAGCAT TTGTATAACT TAGTTGGTCA AAAGATGCTT TTGCATATGC	3300

55

	TTTCATAAGT GATGCTTTAT TAGCAAGAAT ATGTGTTTCGC AGAAATTTGT TCTGCATTCT	3420
	ACTTCTACGC TAGTCAATCA GACAATTTTA CCAATCCCCA CTTTCGCGTT TCAAATCAAA	3480
5	CAATACGTCG CTCCTTTCTT CTTATATAAC AATTCTTCTA ACATGATATG TTACTATTGA	3540
	ATTACTGAAC CTGAGTTAGT TATAATCTAA CTTATATTGA AAAGAGATGA GCGTAAGAT	3600
	ATGTTTTTAT GTAAAAGACA AATTGATATC AATGCACGAT TTGGTTTGCC TAGAATTGCA	3660
10	TTTATGAGTG CAGTTGCAAC CATCATTATG TTTTATAGTA GTTATGAAGT AATGTATTTT	3720
	TTATCTAATA CGCCATTATC AGATAGACAT TTTCTCATCT TTTTATTACT TGTATTTATG	3780
	ACGTATCCAT TACATAAAAG TATACATTTA TTATTTTTCT TACCATATAG AAAATCGTTT	3840
15	AAAGTTCATA AGTTAACTAA AAGAAAATGG CTTATATTCT ATAATACCTA CGTCAATCAA	3900
	CCTGTACACA AATTTTATTT TTGCATTAAC TTAATATTGC CGTTAATTAT CTTATCTGCA	3960
20	ATGTTTCGTTT ATCTAACAAT TTCATTCCCG CAATATGGAC ATTATTTTAT GTTCTTATTG	4020
	GCATTGAATT TCGGTATTTT CATTACAGAT TTATTATATT TAAAAATAAT TATATTTTCT	4080
	AATTATGGAC AATATATAGA AGAACATAGT ACAGGTATTA ATATTTTGAA AAAAATTAAA	4140
25	AATCCATATC ATTTATAACA AAATAATTAT AGCAAGGTGT TATTATTGT TTTTAGGCTA	4200
	TGTAATAgcT tACAATCAAA TGTATATAGA CCTTGTTTTT TTATTTTCAT CAATTTCTAC	4260
	CCCTAAACCT AATGCTCTAG TCTGATGTCA TGGGTTATTG ATTGGTGATA ATATAAACT	4320
30	ATGTTATATT CACGATGATT AACTTACAAA GGAGTTTCAA CTATGAAGAT GATAAACAAA	4380
	TTAATCGTTC CGGTAACAGC TAGTGCTTTA TTATTAGGCG CTTGTGGCgC TAGTGCCACA	4440
	GA CTCTAAAG AAAATACATT AATTTCTTCT AAAGCTGGAG ACGTAACAGT TGCAGATACA	4500
35	ATGAAAAAAA TCGGTAAAGA TCAAATTGCA AATGCATCAT TTA CTGAAAT GTTAAATAAA	4560
	ATTTTAGCTG ATAAATATAA AAATAAAGTT AATGATAAGA AGATTGACGA ACAAATTGAA	4620
	AAAATGCAAA AGCAATACGG CGGTAAAGAT AAATTTGAAA AGGCCCTTCA ACAGCAAGGT	4680
40	TTAACAGCCG ATAAATATAA AGAAAATTTA CGTACTGCTG CTTATCATAA AGAATTACTA	4740
	TCAGATAAAA TTAAATCTC TGATTCTGAA ATTAAAGAAG ACAGCArGAA AGCTTCACAC	4800
45	ATTTTAATTA AAGTTAAATC TAAGAAAAGC GACmAGAAG GCTTAGATGA TAAAGAAGCG	4860
	AAACAAAAAG CTGAAGAAAT TCAAAAAGAA GTTTCAAAAG ATCCAAGTAA ATTTGGTGAA	4920
	ATCGCTAAAA AAGAATCAAT GGATACTGGT TCAGCTAAAA AAGATGGCGA ATTAGGTTAT	4980
50	GTTCTTAAAG GACAACTGA TAAAGATTTT GAAAAAGCAC TATTTAAGCT TAAAGATGGT	5040
	GAAGTATCAG AGGTTGTAA ATCAAGCTTT GGATATCATA TTATTAAAGC TGATAAACCA	5100

55

AAAAATCCAA AATTATTGAC TGATGCATAC AAAGATCTAT TAAAAGAATA CGATGTTGAC 5220
TTTAAAGATC GTGATATTAA ATCAGTTGTC GAAGATAAAA TCTTAAACCC TGAAAACTT 5280
5 AAACAAGGTG GCGCACAAAG CGGACAATCC GGCATGAGCC AATAACACAA AACCGAGCGA 5340
CCGTGGTTCA AAAATCATA CACGGCCGCT CGGTTTTTTC GCATTAAAAA TCGGACAGAT 5400
GAGCTCATGT TTCAGTATAC TCATCTGTCC GATATCTTTT AATTCTTAAT CGAGTGATTC 5460
10 AGGATTGTAG AATCTACGAT TTTCAAGACC AAATATTTTA TCTGTAAACT GACCCTTGTC 5520
AGTTTTTTTA TATGCCTTTT CAAACATATT CATTCTAGCA TCGATATTAT CGATATAGCA 5580
TAAAAATTTCT GCTTCTTTTA AGTATGGCAG TTTTGGAGAA CCATACTCTA ACTTACCATG 5640
15 ATGAGATAAA ATCATATGTC TTAACAACAT GATTTCTTCT CCTTCAATGT TCAATTCACG 5700
AGCTGCTTCA ACTACTTCAT CACTCGCAAT CGAGATGTGT CCTAATAAGT TACCTTCGAC 5760
TGTATACGAC GTCGCAACAG GACCACTCAA TTCTCTAACT TTACCAATAT CATGCAAAAT 5820
AATACCACTA TATAACAAAC TTTTGTTTAA CAATGGATAA ATGTCAAAA TTGATTTTGC 5880
AATACGTAAC ATCGTTAATA CATGATAGCT TAAGCCACTC GCAAAGTTAT GaTGATGAGA 5940
25 ACTAGCAGCT GGATATGTGT AAAATCGTTC TTGATATTTT TTCAATAAAT GACGTGTGAT 6000
ACGTTGTAAA TTAGCATTTT CAATATCTAG CAAATAATGA GAAATCTCTT CTTGTATTTT 6060
TGCCGGTGAT AAAGGTGCAC CATCTACAAA TTGTTCTGTT TTTAATTGAT CTTCAAGTTG 6120
30 CGCTAGTCTA ATTTGGTTGA CTTTCATCTG TTTATTTCCG CGATAGTTTA TGATGTCACC 6180
TTTAACATGT ACAATTTCTT CAGGCTTGAT TGTTGCCATA TCATTTTTTG TAGCCGTCCA 6240
AAATTTGCT TCAATTTTAC CACTTTTATC TTGCAAATGT AATGTCATAT AATCTTTACC 6300
35 TTGTGCTGTT ACACCCTGTG TAGCTTTATG CACTAAGAAA AAGTGATCAA CTGAATCTCC 6360
GGGA~~T~~TTAGA TTCTCTATAT TTCTCATCGT TTCCCGCCTT CCTCTATTTT GTTTAATGTA 6420
ATCACTTCTT TTGATGGAAC AATATTATCT TTTACACATG TAAAGTATAG TACTTGATAG 6480
40 TGTTCTGATA ATGATCGTAA ATAATTCAAC ATTTTTTCAG TACGTTTTTT ATCAAAATGA 6540
ACAAATGCAT CATCAACAAT TAATGGGAAC GGATAATATG GTCTTAGTAC CTTAATTAAA 6600
CTGATACGTA AAGCTACATA AAGTAATTCT TTTGTAGATT GACTTAGTTC AACAGGATCA 6660
45 TATAATTGAC CATTAACATG TTTAACCGTA ATTGAATCTT CATTATAGTT AATCATCGTA 6720
TATCTGCCAT CTGTTAAATG CTTCAATATT TCTACCGCTT CATTAATAAC TTGAGGCAAA 6780
50 CGTTTATCTT TAATTTGTTT AATGTGTTCA TCAACTAAAC TTTGTAAATA ACTTAACTT 6840
GCCCAATCTT TTGCGATATC ATTAAGTTGA TTTTAAAGAC TGTGATATTC ATGTCTTAAA 6900

55

GCTTGCATTT CAAGATATTG CTCATTATAT TCGTCAACTT GAGTAGCCAA TAAATGATCT 7020
 TCTTCTTCAA GTTGTGCAGT TGTMTTTC CACTAACTAG AACTTAATTC ATAAGAATAG 7080
 5 TTTTGGTTCT CAAGATATTT AGTTAAATCA TTAAAACGAC TCAAATTACT AGTATAAGTT 7140
 TGGTAATCTT CATGATGTTG GTAAAAATCT TCTTCAGTAC CAACATTGAT AAAATCGAAT 7200
 AGTGCTGTAA TTTCTTTATT ATTTTCTTCT AATTGAGCAT TTAAATGATT TAATTCATTT 7260
 10 GTAACAAGTT TGGTATTTTC AGCATTAAATA CGCCATTTT CATTCGTGTC TTCAGCTGAT 7320
 TTCAACCATT GTtGCACATC GTGGAATAAA GATAATTTGT TGAAATAAAC AAATTGTGAT 7380
 TTTGTAACAG CTTCAGCATG ATTGTAGAAT GTATCTAATT CTTGAACCAA TTGCTGGCGT 7440
 15 TGTGATTTA AATCACTGAT ATGTTGATCT AATGCTTTAA TATTGCGCAT TGTAGAAATA 7500
 CTATCAACAA TTAAATCATT TGAAATTTTA GATGATAAGT ATAATTCATC CTTAACGTTT 7560
 TCAACTGTCG ATTGTAATTC ATCATGACGC CCTTTCGCAT CATTTAAACG ACCTTCAATA 7620
 20 TACTGACGTT TCTCTTCTAA AATATCTTTA TTTTTCAAAG CTTGTTGCCA GTGATCACGA 7680
 ATGCGATATT GTCATCAAG ATCAAAATCT AAGTCATAAT TTTCATCTAA AATGGCTAGT 7740
 25 TGTGCTTTAA TTTCTTCGAT TTCATCTGTG ATGGCCTCGC TATAATCTAC TTCTTTTGAT 7800
 TTAGACATGA TGATACCGAT AACAAATACT AAAGTTAATA CTGCGAAAAT AATACCAAAC 7860
 AACATGTTGT TTGAAATAAA TGAGAAGGCA GTTAAACCAA TACCTACTAA TGTTAAAAGr 7920
 30 ATAAACGTTG TTCGKAACAA TTTTGTGACGT TTTTGTtTTT CTT 7963

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:
 35 (A) LENGTH: 3958 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

ATATTGTCTT TACAATAGTT TGCTATGGAG GTAATTAACC AATAGGAGGA ATTTATAATG 60
 45 GCAGTAATTT CAATGAAACA ATTACTAGAA GCGGGTGTTT mCtTCGGTCA CCAAACACGT 120
 CGTTGGAACC CAAAAATGAA AAAATATATC TTCACTGAGA GAAATGGTAT TTATATCATC 180
 GACTTACAAA AAACAGTGAA AAAAGTAGAC GAGGCATACA ACTTCTTGAA ACAAGTTTCA 240
 50 GAAGaTGGTG GACAAGTCTT ATTCGTAGGA nCTAAAAAAC AAGCACAAGA ATCAGTTAAA 300
 TCTGAAGCAG AACGTGCTGG TCAATTCTAC ATTAACCAA GATGGTTAGG TGGATTATTA 360

55

	GAAGATGGTT TATTCGAAGT ATTACCTAAA AAAGAAGTAG TAGAACTTAA AAAAGAATAC	480
	GACCGTTTAA TCAAAATTCCT AGGCGGAATT CGTGATATGA AATCAATGCC TCAAGCATT	540
5	TTCGTAGTTG ACCCACGTAA AGAGCGTAAT GCAATTGCTG AAGCTCGTAA ATTAAATATT	600
	CCTATCGTAG GTATCGTTGA CACTAACTGT GATCCTGACG AAATTGACTA CGTTATCCCA	660
	GCAAACGACG ATGCTATCCG TCGGTTAA TTATTAATCTG CTAAAATGGC AGATGCAATC	720
10	TTAGAAGGTC AACAAAGCGT TTCTAATGAA GAAGTAGCTG CAGAACAAAA CATCGATTTA	780
	GATGAAAAAG AAAAATCAGA AGAAACAGAA GCAACTGAAG AATAATCAAC TGTGAATCT	840
	GACTTAGATA TAGTTTAAAT GGGTGATAAG ATATTAATGC TTATCACCTT TTTTAAAAAG	900
15	AAAATCGAGG CAAATTACAA ATATTCAATT AGAGTATTGG CAATCTTGCC TATAATAATG	960
	CTAAAATCAT AATATATAAN ATGATAACTT ATTGGAGGAA TAATGAATGG CAACTATTTC	1020
20	AGCAAACTT GTTAAAGAAT TACGTGAAAA AACTGGCGCG GGTATGATGG ATTGTAAAAA	1080
	AGCGCTAACT GAAACTGATG GTGACATCGA TAAAGCGATT GACTACCTAC GTGAAAAAGG	1140
	TATTGCTAAA GCAGCTAAAA AAGCAGACCG TATTGCGGCT GAAGGTTTAG TACATGTAGA	1200
25	AACTAAAGGT AACGACGCAT TATCGTTGAA ATCAACTCTG AACAGACTT TGTGCTCGT	1260
	AACGAAGGTT TCCAAGAGTT AGTTAAAGAA ATCGCTAATC AAGTATTAGA TACAAAAGCT	1320
	GAAACTGTTG AAGCTTTAAT GGAAACAACT TTACCAAATG GTAAATCAGT TGATGAAAGA	1380
30	ATTAAAGAAG CAATTTCAAC AATCGGTGAA AAATTAAGTG TTCGTCGTTT TGCTATCAGA	1440
	ACTAAACTG ATAACGATGC TTTCGGCGCT TACTTACACA TGGGTGGACG CATTGGTGTA	1500
	TTAACAGTTG TTGAAGGTTT AACTGACGAA GAAGCAGCAA GAGACGTTGC TATGCATATC	1560
35	GCTGCAATCA ACCCTAAATA TGTTCCTTCT GAACAAGTTA GCGAAGAAGA AATCAACCAC	1620
	GAAAGAGAAG TTTTAAAACA ACAAGCATT AATGAAGGTA AACCAGAAAA CATCGTTGAA	1680
40	AAAATGGTGG AAGGACGTTT ACGTAAATAC TTACAAGAAA TTTGTGCTGT AGATCAAGmT	1740
	TCGTTAAAAA CCCTGATGTA ACAGTTGAAG CTTTCTTAAA AACAAAAGGT GGAAAACCTG	1800
	TTGACTTCGT ACGCTATGAA GTAGCGGAAG GTATGGAAAA ACGCGAAGAA AACTTTGCGG	1860
45	ATGAAGTTAA AGGACAAATG AAATAATCTG TCATAAAGTA AAACAAGGAA GAAGACACCT	1920
	TTAATGTTGC TTTATTAAAA TGTAAATCAT TCTAATAAAA CGACAACTGT GTCTTCTTTA	1980
	CTTGATATG TTACATATAT TCACGATAGA GAGGATAAGA AAATGGCTCA AATTTCTAAA	2040
50	TATAAACGTG TAGTTTTGAA ACTAAGTGGT GAAGCGTTAG CTGGAGAAAA AGGATTTGGC	2100
	ATAAATCCAG TAATTATTAA AAGTGTGCT GAGCAAGTGG CTGAAGTTGC TAAAATGGAC	2160

55

	TTAGGTATGG ACCGTGGAAC TGCTGATTAC ATGGGTATGC TTGCAACTGT AATGAATGCC	2280
	TTAGCATTAC AAGATAGTTT AGAACAATTG GATTGTGATA CACGAGTATT AACATCTATT	2340
5	GAAATGAAGC AAGTGGCTGA ACCTTATATT CGTCGTCGTG CAATTAGACA CTTAGAAAAG	2400
	AAACGCGTAG TTATTTTTCG TGCAGGTATT GGAAACCCAT ACTTCTCTAC AGATACTACA	2460
	GCGGCATTAC GTGCTGCAGA AGTTGAAGCA GATGTTATTT TAATGGGCAA AAATAATGTA	2520
10	GATGGTGTAT ATTCTGCAGA TCCTAAAGTA AACAAAGATG CGGTAAAATA TGAACATTTA	2580
	ACGCATATTC AAATGCTTCA AGAAGGTTTA CAAGTAATGG ATTCAACAGC ATCCTCATTC	2640
	TGTATGGATA ATAACATTCC GTTAACTGTT TTCTCTATTA TGGAAGAAGG AAATATTAAA	2700
15	CGTGCTGTTA TGGGTGAAAA GATAGGTACG TTAATTACAA AATAAATTTA GAGGTGTAAA	2760
	ATAATGAGTG ACATTATTAA TGAACTAAA TCAAGAATGC AAAAATCAAT CGAAAGCTTA	2820
20	TCACGTGAAT TAGCTAACAT CAGTGCAGGA AGAGCTAATT CAAATTTATT AAACGGCGTA	2880
	ACAGTTGATT ACTATGGTGC ACCAACACCT GTACAACAAT TAGCAAGCAT CAATGTTCCA	2940
	GAAGCACGTT TACTTGTTAT TTCTCCATAC GACAAAACCT CTGTAGCTGA CATCGAAAAA	3000
25	GCGATAATAG CAGCTAACTT AGGTGTTAAC CCAACAAGTG ATGGTGAAGT GATACGTATT	3060
	GCTGTACCTG CCTTAACAGA AGAACGTAGA AAAGAGCGCG TTAAAGATGT TAAGAAAATT	3120
	GGTGAAGAAG CTAAAGTATC TGTTGAAAT ATTGTCGTG ATATGAATGA TCAGTTGAAA	3180
30	AAAGATGAAA AAAATGGCGA CATTACTGAA GATGAGTTGA GAAGTGGCAC TGAAGATGTT	3240
	CAGAAAGCAA CAGACAATTC AATAAAAGAA ATTGATCAAA TGATTGCTGA TAAAGAAAAA	3300
	GATATTATGT CAGTATAAAA CTAATATACA ATGACATATT AAAATGCCAG TATTAAACGA	3360
35	TAATGTAACA TTTAAAATGG GCATGTTTAA TTAAATCAAA GATGCATGTG ATAATTTAAA	3420
	TTCAGAATGA GCATAAAAAT GGTGTTTAAA CAAGTTAATT AAACATATAC TTTATAAATA	3480
40	ATAGGCATTA GGTATATTGC TATAATAAAG TTATGTAATT TTAAACCTCA GTATGTATGT	3540
	CACATTTCTG GTGTAAACTG TACCGAGTCA GACTTTGGTA CAGTTTTTTT ATTTGCTTAT	3600
	TCAATGCATT AAATGAGTAT GATAAAATGA TAATGATTGT TTAGTAACTT ATACTATATG	3660
45	ACAGAGATGA TCAGGCTCGG AGGAAAGACC ATGTTTAAAA AGCTAATAAA TAAAAAGAAC	3720
	ACTATAAATA ATTATAATGA AGAATTAGAC TCGTCTAATA TACCTGAACA TATCGCTATT	3780
	ATTATGGATG GTAATGGGCG ATGGGCTAAG AAGCGAAAAA TGCCTAGAAT TAAAGGTCAT	3840
50	TACGAAGtAT GCAAACAATA AAAAAATTA CTAGGGTAGC TAGTGATATT GGTGTTAAGT	3900
	ACTTAACTTT ATACGCCTTT TCCACTGAAA ATTGGTCAAG ACCTGAAAGT GAAGTAAA	3958

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5333 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

10 ATTAAACAA CTTAATATAC CTATTTATGG TGGTCCTTTA GCATTAGGTT TAATCCGTAA 60
 TAAACTTGAA GAACATCATT TATTACGTAC TGCTAAACTA AATGAAATCA ATGAGGACAG 120
 15 TGTGATTAAA TCTAAGCACT TTACGATTTC TTTCTACTTA ACTACACATA GTATTCCTGA 180
 AACTTATGGC GTCATCGTAG ATACACCTGA AGGAAAAGTA GTTCATACCG GTGACTTTAA 240
 ATTTGATTTT ACACCTGTAG GCAAACCAGC AAACATTGCT AAAATGGCTC AATTAGGCGA 300
 20 AGAAGGCGTT CTATGTTTAC TTTCAGACTC AACAAATTCA CTTGTGCCTG ATTTTACTTT 360
 AAGCGAACGT GAAGTTGGTC AAAACGTAGA TAAGATCTTC CGTAATTGTA AAGGTCGTAT 420
 TATATTTGCT ACCTTCGCTT CTAATATTTA CCGAGTTCAA CAAGCAGTTG AAGCTGCTAT 480
 25 CAAAAATAAC CGTAAATTG TTACGTTCCG TCGTTCGATG GAAAACAATA TTAAATAGG 540
 TATGGAACCT GGTATATTA AAGCACCACC TGAAACATTT ATTGAACCTA ATAAATTA 600
 TACCGTACCG AAGCATGAGT TATTGATACT ATGTACTGGT TCACAAGGTG AACCAATGGC 660
 30 AGCATTATCT AGAATTGCTA ATGGTACTCA TAAGCAAATT AAAATTATAC CTGAAGATAC 720
 CGTTGTATTT AGTTCATCAC CTATCCCAGG TAATACAAAA AGTATTAACA GAACTATTAA 780
 TTCCTTGAT AAAGCTGGTG CAGATGTTAT CCATAGCAAG ATTTCTAACA TCCATACTTC 840
 35 AGGGCATGGT TCTCAAGGTG ATCAACAATT AATGCTTCGA TTAATCAAGC CGAAATATTT 900
 CTTAECTATT CATGGTGAAT ACCGTATGTT AAAAGCACAT GGTGAGACTG GTGTTGAATG 960
 CGGCGTTGAA GAAGATAATG TCTTCATCTT TGATATTGGA GATGTCTTAG CTTTAACACA 1020
 40 CGATTGAGCA CGTAAAGCTG GTCGCATTCC ATCTGGTAAT GTACTTGTTG ATGGTAGTGG 1080
 TATCGGTGAT ATCGGTAATG TTGTAATAAG AGACCGTAAG CTATTATCTG AAGAAGGTTT 1140
 45 AGTTATCGTT GTTGTTAGTA TTGATTTTAA TACAAATAAA TTACTTTCTG GTCCAGACAT 1200
 TATTTCTCGA GGATTTGTAT ATATGAGGGA ATCAGGTCAA TTAATTTATG ATGCACAACG 1260
 CAAATCAAA ACTGATGTTA TTAGTAAGTT AAATCAAAAT AAAGATATTC AATGGCATCA 1320
 50 GATTAAATCT TCTATCATTG AAACATTACA ACCTTATTTA TTTGAAAAA CAGCTAGAAA 1380
 ACCAATGATT TTACCAGTCA TTATGAAGGT AAACGAACAA AAAGAATCAA ACAATAAATA 1440

55

	GCTTTTCTT TATATATGAT GAGCTTGAGA CATAAATCAA TGTTCATGTC TCTACAAAGT	1560
	TATATTGGCA GTAGTTGACT GAACGAAAAT GCGCTTGTA CAAGCTTTTT TCAATTCTAG	1620
5	TCAGGGGCCC CAACATAGAG AATTCGAAA AGAAATTCTA CAGGCAATGC GAGTTGGGGT	1680
	GTGGGCCCCA ACAAAGAGAA ATTGGATTCC CAATTTCTAC AGACAATGTA AGTTGGGGTG	1740
	GGACGACGAA ATAAATTTTG AGAAAATATC ATTTCTGTCC CACTCCCGAT TATCTCGTCG	1800
10	CAATATTTTT TTCAAAGCGA TTAAATCAT TATCATGTCC AATCATGATT AAAATATCAC	1860
	CTATTTCTAA ATTAATATTT GGATTGGTG AAATGATGAA CTCTTGCCT CGTTTAAITG	1920
15	CAATAATGTT AATCCATAT TGTGCTCTTA TATCTAAATC AATGATAGAC TGCCCCGCCA	1980
	TCTTTTCAGT TGCTTTCAAT TCTACAATAG AATGCTCGTC TGCCAACTCA AGATAATCAA	2040
	GTACACTTGC ACTCGCAACA TTATGCGCAA TACGTCTACC CATATCACGC TCAGGGTGCA	2100
20	CAACCGTATC TGCTCCAATT TTATTTAAAA TCTTTGCATG ATAATCATT TGTGCTTTAG	2160
	CAGTTACTTT TTTTACACCT AACTCTTTTA AAATTAAAGT CGTCAACGTA CTTGATTGAA	2220
	TATTTTCACC AATTGCCACA ATGACATGAT CAAAGTTACG GATACCTAAA CTTTTCATAA	2280
25	CTGCTTCATC TGTAGTGTCT GCAACAACCG CATGAGTAGC GATATCACTA TATTCATTCA	2340
	CTCTATTTTC ATCATGGTCG ATGGCCATTA CATCCATGTC TAATGCATT CAACTCACGAA	2400
	CGATACTACC TCCAAAACGA CCTAGACCGA TGACTACATA TTCTTTACCC ATACTCGCCC	2460
30	TCCATTAAAT GATTTTCATC AATTCATTGA AAATATAAAT TTAAAATTAT TATAAATGAG	2520
	TACCCCAACT AAATTATCTA AATGCAGTAA TGCAAGTAAA TGAAAGTTGG GGTATCGTCT	2580
	CAACTTATGA TTTCTTCTCT TCAACATATT CTTTGTGCGA AACAAATAAT CTTAATAATA	2640
35	ATATTAACGA TGGAAGTAAT AAAAGTAAAC CTAAAATAAA GACAATCACT AATGTCCAGC	2700
	CCATTTCTGG ATTAACATAT GCATCTGTAA TTTTACAAA CGGATATAAA AGGTATGGCA	2760
40	ATTTACTAAT TCCATAGCCA AAGAACGCGA ACATCATTTG TAAAATAACA AATACAAAAG	2820
	CCAAACCATG TTTTCTCTTA AAGAATGTTA ACAATGAAGC TAATGCAAAG AATAAGAAAC	2880
	TTATACCAAA CATCCACCAA TAGTCAAAAA CAGCTGAATA AAAATGTTCA GAATTTTGAA	2940
45	TGCGTAATGA TAGAAATACG AATAAACAAA TGATAATCAT CGGCGGCCCT AAAAATATGT	3000
	GCCATTGTCT TGTTAAATTA TATGCTGGTT CGTCATTTGC TTTTITAGCA TAATATGTCA	3060
	AAAATCCTGA TGAAATATAT AAAACTGAAA TAATTGCCAA GAATACTACA GACCAAGCAA	3120
50	ATGGGCTTAA TAATAACTGC ACCCAATCTA GATCGATAAC ATTGTTTCGA ACATTAATAT	3180
	AGCCACCTTC TGTAATAGTT AAAGCAGTAG ATAATGAAGC TGGAATTAAT AATCCACTTA	3240

55

	A	A	A	A	A	A	3360
	C	T	A	A	A	A	3420
5	A	T	A	C	A	T	3480
	T	T	T	C	A	C	3540
	C	A	T	A	A	T	3600
10	T	C	A	T	T	T	3660
	T	T	T	A	T	T	3720
	T	A	T	G	C	C	3780
15	G	C	G	C	A	T	3840
	T	T	C	T	A	G	3900
20	A	G	T	A	G	A	3960
	A	A	A	C	A	T	4020
	A	T	G	T	A	A	4080
25	T	A	A	G	A	A	4140
	T	A	C	A	C	C	4200
	T	G	C	T	A	T	4260
30	A	G	A	A	G	T	4320
	G	A	C	A	T	A	4380
	A	A	T	G	C	C	4440
35	G	T	T	A	A	C	4500
	T	G	C	A	G	C	4560
	A	A	T	A	A	T	4620
40	A	T	A	T	A	A	4680
	T	G	G	A	G	T	4740
	A	A	G	T	C	T	4800
45	C	C	A	T	C	T	4860
	A	A	T	G	C	G	4920
50	A	G	T	C	A	T	4980
	T	C	T	T	T	T	5040

55

	GAATTTCAAT GTATAATTGT GTATATTACA TTAGAATAAA GCACGAAGGA GCATGATACA	5160
	TGTCAGAAAT AATCGTTTAT ACGCAGAATG ATTGTCCACC TTGTACATTT GTAAAAAATT	5220
5	ATCTAAATGA GCATCACATT GATTTTGAAG AGAGAAATAT CAACAATCAA CAATATCGAA	5280
	ACGAAATGAT AGATTTTGAT GCTTTTTCAA CTCCGTTTAT TTTGTTGAAT GGC	5333

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11126 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

20	ATACGTGACC CTTTATCCGA AAATTTCTTT TCATATTCTG TTAAATATT ACTGCCATCG	60
	TCTTCTTGAT GTAAATTTAG ATTTATTTTT GTAAAATACA TTCCAAATTG AGACATACTT	120
	TCTAAACTGT AGGCAAATAG TCCTCTGTTA TCAGTTTAA AATGTAAATC TCCTTCATCA	180
25	TTTAAGATTT GTTGATACAA CGCTAAAAAC GTATGATACG TTAAACGTCG TTTTGCATGA	240
	CGATTTTTTG GCCATGGATC TGAAAAGTTC AAATAAATAC GCGAAACTTC GCCGTCTTTA	300
	AAATATTCAT TTAATTCAAT GGCATCATT CAAATAATCT TTAAATTTGT TAAACCCATC	360
30	TCTTTAACTT TATCCAATAC TTTATAAAGC ATACTTTTCT CACGTTCAT TGAAATATAG	420
	TTAATATGAG GATTTTGAGC AGCTAATGTT GTAATAAACT GCCCCATACC CGAACCAATT	480
	TCAATGTGTA TCGGTTGCGT TTTATCAAAC CATTCACTCA TTTTCCCTGc ATGTTGACCG	540
35	TCCATGTCAA CCAATTCAGG ATGATCTTTT AAATAATCTT CAGCCCATGG TTTGTATCGA	600
	ACTCTCATAT TTTATCTCC TCTTAAATAA ACATGTTACT ATTCATAACT TCATTTAGGA	660
	ATTTAAGCCA AGTGTTTATA TCCTTATATC TTTTGTGCTC TTCATACCAT TGAACAAGAC	720
40	CTATAGATTG AATTACCGTA TACCATTTCA TACGTTTATT TAAATTCAAG CTCTCTTGAA	780
	CACCATATGT TTCAAGCCAT TCAGACCATT GTTGTGTGG AACATAGTTG TAAAGCAGCA	840
45	TTCCGATATC AATTGCCGGG TCTGCAATCA TTGCACCTTC CCAATCAACT AAAAATAGTT	900
	CATCTCGATC GGATAATAAC CAATTATTAT GATTACATC ACCATGTACA ACAGTGAAAA	960
	AACGCGAATC TAAACTCGGT ATATGCTCTT CTAAATAGGT TAATGATTTT CTCACAATAT	1020
50	GATGTGTTAA AACTTCTCTT GATAAAGAGG CATTAAATTT ATTAAGCATA ATCTCAGGAG	1080
	TAATAGGTTT CATTTCCATA CGCTTTAACA TACTTAATAA AGGTCTAGAA TTGTGTATCT	1140

	TTTTCCAATG TTGTGCTGTA ACAACCTCGC CTGTTTCTAT GCGTTTCGTC CATACTAATT	1260
	TGGGCACAAT ACCTTCTGCT GATAATGCCG CAATAAATGG ATTTGAATTT CGTTTTAAAA	1320
5	ACAACTTTTG TCCATCTTGT TCAGCCATAT ATGCTTCACC AGATGCACCA CCTGCTGAAT	1380
	CAAGTGTCCTA CCCTAATTGA TAAACTGCT CCAACTCGTC CACCTCACTT TCAATTAGAA	1440
	AATGGCTCTA GAAATAGGTT TTTCAAGAGC CATATATTCT AATTTATAAC ACCATACTGG	1500
10	TACAAATATT ATGTCCAGAT AATTATTGTA AATCCTCAAC CAATGCCTAC ATTACACGAC	1560
	TAAATTTAAA TCGTAATGTC TGTCAATTGAC ACCATACATT CTATAGTCAC TTACTTGACA	1620
	TATAATGTTA CCGTGTCTAA AACTACATGT TTTTGAATCT CTGTAGGCGA TAAACTcTAG	1680
15	TTTTCAAAAT AATTGCTATC CCATTTTCAT GGTTAGCATA AATTTATGAA CTGTAACATT	1740
	TACGTACTTA GTAAAATATG ATGCACATCA TATTTGTtAC TCATAGAAAA TTTTATAAtT	1800
20	TTTATCATTa TATTTCAACT GAAAATGAGA AACAAAATGG CACTTTTtAC TAATATGTGT	1860
	TTTCTAAACA ACACTTTTTAA GCTTCGTTTT AAATTATAAC ATAATTCACt TACGAAAGTT	1920
	GATAAATTTA AGTAATTTAA TCTAAAAATA TGATGAAAGA ATTTTAAATA CTGTGTGACT	1980
25	CTATATACTT TTCAAATCCT TCTTGTAGTT GACGTGTAAT TGGGCCAACT TTACCATCAT	2040
	TAAGTGGTTC ACCATCTAAT TTAATAACAG GTGTAACCTC AGCTGAAGTA CTTGAAACAA	2100
	TAAGTTCATC TCGTTTTTTC AAGAAATCTA CAGTAAACGT TTCTTCTTTA AATGGGATGT	2160
30	TATAGTCTTC GGCAATTTTT TTAATTACAA TTCGTGTAAT ACCATTAAGA ATATAGTTGT	2220
	TAATCGGATG TGTATAAATC ACACCGTCTT TAATTGCATA AGCATTACTT GAAGATCCTT	2280
	CAGTTACAGT TcCACCTCGA TGTTGAATTG CTTCAACTGC ATTATATTTc ACAGCATATT	2340
35	CTTTTGCTAA TACATTcTCC TAATAAGTTC AAGCTTTTAA TGTCGCAACG TAACCATCGG	2400
	ATATcTTCAA CGGTAACACC ATTCACACCA TTTTCTAAAT GATCATAAGG ACGATCATAA	2460
40	CTCTTTGTAT AAGCAACAAT TGCTGGTTCT ACTTCAGGTG TCGGGAAGCT ATGATTCTCT	2520
	TCAGCTACAC CACGCGTTGC TyGAATATAA ATTGCCCCAG TTTCAATTTG ATTCATATCA	2580
	ACTAATTTAC GAGATAGTTC AATTAATTCT TCTACAGAAT AATTtAAATC TAAACCAATC	2640
45	TCATTGGCAC TACGTWAAAw TCTTTCATAA TGTTCTGTTA CTGTAAATAA CTTACCATTA	2700
	TATACTCGAA TGTATTcATA AATACCATCG CCAAATACGT ATCCTCTGTC GTTGTATGAA	2760
	ACCTTTGCTT CACTTGGACT TACAACTCA CCATTTAAAA AAATTTTTTTC CATATATTAT	2820
50	TCCTCCACGC ATAATGAATA AATTGCTTCT AAGTAAATAC TAGTTGCGTT AAATAACTGT	2880
	TTTTTAGTGA TATATTcATT TTTCTGATGC ATTAAATCTT CAGAATCACT AAACATTGCG	2940
55		

	TCAGTCATAT	CATTGTGTTG	ATTTCTATAT	GCAGTAACTA	ACTTTTGTAC	AAAAGGATCA	3060
	TTTTTATCAA	CATAATGTGG	TGGTTGGACT	TTACCTAATT	TCACCTTCAA	GCCATATTGT	3120
5	TGAATCTCAT	TTGCAAAACG	ATCCATAGCT	TTTTCAAATT	CAAATCCTTC	TGGGTAGCGT	3180
	AAGTTGATAC	CGAAAAGACC	TGCGTTTTCA	TTATCATATG	TAATAACACC	AATGTTAGTT	3240
	GTCACGTCAC	CCATGACATC	TGTATGGAAT	TTCATTCCCA	TCTTTTCACC	AAAATCTGAA	3300
10	TTAAATAAGT	AGCGATTACT	AAATGCTACA	AACGCTTG TG	CATTATTATC	AAGATTTAAT	3360
	GATGCTAAGA	ATTTTAGTAA	GTAAAGACCC	GCATTCACAC	CGATAGATGG	ATCCATACCA	3420
	TGAACCGCTT	TACCTTCAAC	TGTTAAAACT	AGAATGCCAC	TATCAACAGT	ACTATCACCT	3480
15	TGTAAATGAT	TTTGTCTCTA	AAAGTACTCA	AAGTCTTGAA	TAACATCTGT	CATATTTTCT	3540
	TTAACAAGCA	CTCTTGCTTC	TGCATGATCA	GGTACCATGT	TGTAACGTTC	ACCAGATTTA	3600
20	AAAGTTATTA	ATTCATAATC	AGGTTTCATCT	TGATCTTCAG	TAAGTTTATT	TTGAACTAAA	3660
	TCAAATGTTG	TAATGCCTTT	TTCACCATGA	ATACATGGAA	ATTCTGCATC	TGGTGCAAAA	3720
	CCTAATGTTG	GCATTTCTTC	TGTTTTAAAA	TAGCGATCCG	TACATTTCCA	ATCAGATTCT	3780
25	TCATCCGTAC	CAATAATCAT	ATGAATACGT	TTCTTCCAAT	CCACATTCAT	ATCTTCTAAT	3840
	ATCTTAATTG	CATAATAAGC	AGCAATTGTT	GGACCTTTGT	CATCAAGTGT	ACCTCTAGCT	3900
	ATGATAGCAT	CTTCTGTTAC	AACCGGCTCG	AACGGATTAC	TATCCCATCC	ATCACCAGCA	3960
30	GGAACAACGT	CAACATGACA	TAAGATACCT	AATACGTCAT	TTCTTTTACC	TGCCTCAATT	4020
	CTTCTGCAA	TATGATCCAC	ATCATGTGTT	GTAAATCCAT	CTCTATGTGC	AATTTTCATAC	4080
	ATGTAGTCTA	ATGCCTTACG	AGGACCTGGA	CCAACTGGTG	CGTCTTCTGA	TGCTTTTGCA	4140
35	TCATCTCTCA	CACTTTCAAT	TGCTAATAAT	CCTTTTAAGT	CATTAATGAT	TTGATCTTCG	4200
	TATGTGTTGAA	CTTTTTCTTT	CCACATTCGA	AATCGACTTC	CTTTTTTCTA	TAAGTTAAAT	4260
40	TCTATTTTAC	ATGAAAAGAT	ATAAAAAC TA	CAATAAGATG	TCAGAAAATA	ATAAAAAGGA	4320
	ACAAAACGAT	GCTATTGATA	TGACACAAAT	CATAAATAGC	TGCTTTGTTC	CTTTTTTAAT	4380
	TTATATATTT	AAAATACACA	TATTCAAGAG	CTCGAGATAT	AAGTCAATGT	ACTAGGCACA	4440
45	CAATTTAATA	TTGACAGTAA	TTAACCGAAC	GAAAATGCGC	CCCGGGGCCC	CAACATAGAG	4500
	AATTTTCGAAA	AGAAATTCTA	CAGACAATGC	AAGTTGGCGG	GGCCCCAACA	TAGAAGCTGG	4560
	CCAATAGTTA	GCTTTCAATA	ATGTGCAAGT	TGGGGTAAGG	GCCCCAACAC	AGAAGCTGGC	4620
50	CAATAGTCAG	CTTTCAATAA	TGTGCAAGTT	GGGGTAAGGG	CCCCAACACA	GAGAATTTTCG	4680
	AAAAGAAATT	CTACAGACAA	TGCAAGTTGG	CGGGGCCCCA	ACACAGAAGC	TGGCCAATAG	4740

55

	TAAAGAAATA CGTTTTCTTT AGATATTAGT ATTTCTTATG AATGAGTTTC ACGCATGTAT	4860
	TCTTCTTTCT ATATGCATAT TAGCTATGAC TAACGATAAA GAACCTGAAA CACTAATAAA	4920
5	TGTCCTATAG TTTACAATAT TATATTGGCA GTAGTTGACT GAATGAAAAT ACGCTTGTA	4980
	CAAGCTTTTT TCAATTCTAG TCAACCTTGC CGGGGTGGGA CGACGAAATA AATTTTGCTA	5040
	AAATATGATT TCTGTCCAC TCCCTTATCA TTTCTGTCT ACTCACATCT TATTCTTTAT	5100
10	CAGATAATGC ATTTTTATTCT TTTTAAAT CTTCTTCAGT GACGATACGT AAATTATTAT	5160
	TTGGTGTGCG CCACCTTCAT CATCAAATTT ACCTTTTCA ATACTTTCGT CAGTCTTATT	5220
	GTCATATTCG GTAAATTTTG ATTTTCTTC TTTGAAAAAT GCTTTTGGAT TATTTTTTAA	5280
15	TCTATTAGCA TATTCTTTCG GATTGTTTT TACTTCTTTA ATTGTTTCAT TAGCAATTGT	5340
	TCCTAATTGC GTCGCTTTAT CCTTAGCATT ATCTTTATAG CTTTGAGGAT CTTGTTTATA	5400
	TTTATTATAT TCCGTGCTTC AGCTGTGCAC GACTATCTTT ACGTGTAACA AGTACAGCTG	5460
20	CTACAGCGCC ACCTATACCT AAAATCGCTT TAAATAAATT ACCTTTTGCC ATATCAATCG	5520
	TCTCCCTTTT ATTTATAATT TAATTGTCA AAATCATTTT CAGTTAATAA ACGATATTCT	5580
25	CCTGAATCTA AATTGCTGTC CAATTCTAAA TCAGCAATTT TGATACGTCT TAAATGTAAT	5640
	ACCTCATTTT GAATGCTATG AAACATTCGT TTAACCTGAT GATATTTTCC TTCATAAATT	5700
	GTTACGTGTG ACGTTTGATT ATCAATATAA GTTAATATTG CAGGCTTAAC CTTGCCATCA	5760
30	GTCAGTGTGA CACCCTCTTT AAAAGCTTGA ATGTCGTCTT CAGTGATAGG ATTTGCTGAA	5820
	ATAAATTCAT ATTTTTTAGA AACATGTTTG TTTGGACTCA TTAATTCATG ATTAATAATCA	5880
	CCATCATTCG TTATCAATAA AAGCCCTTCT GTATCTTTAT CAAGACGACC AACCGGAAAA	5940
35	ATATTTAGAT GTTGGTATTC AGGTATTAAA TCAATAACGG TTTTGAATG ATGATCTTCA	6000
	GTTGCTGATA TATAACCTTT TGGCTTATTT AACATAATAT AGACATTTTC AATGTATTCT	6060
	ATTAAATCTC CACGAACTGT TATCTTATCG TTTTCTGGTT CTATATGTGT TTTTGGTGAT	6120
40	TTAATTACTT GTTCGTTGAC ATTTACAAGG CCTTTTTTAA GTAACGTGTT GACCTCATTA	6180
	CGTGTACCGA CGCCCATATT TGCTAAAAAT TTATCTATTC TCATCGTAAA AACCTAACTC	6240
	TACGTCTTAA TTTTTCAGGA ATTTACCTA AGAATTCGTC CGCAAGACGC GTTTTAATTG	6300
45	TGATTGTACC GTAAATTAGA ATACCTACTG TAACACCTAA AATAATAATG ATTAAGTAAC	6360
	CAAGTTTAGT AGGTTCTAAG AATAGATTTG CAAGGAAAAA TACTAATTCT ACACCTAGCA	6420
50	TCATAATAAA TGAATACAAG AATATTTTTG CAAAATGAAT CCAACTATAG CTGAATTTAA	6480
	ACTTCGCATA TTTTAAAGA ATATAGAAAT TACATCCAAT TGCAAATAAT AATGCGATAC	6540

55

	ACTTGATAAC TACAGAAGCT AAAATAACAT AAAGTGTAA TTTCTGTTA TCTATACCTT	6660
	GTAACATTGA TGCCGTTACA CTTAATAGTG AAATTAGTAT TGCTACAGGC GCATAATAGA	6720
5	ATAATAAGCG ACTACCATCA TGGTTAGGGT CATGACCTAA AACAATTGGA TCGTAACCAT	6780
	AGAAAACTGT GAATAATGGT TGTGCCAAGG CCATAATTCC AATACTAGCT GGAACAGTTA	6840
	TAAACATTAA TACACCAATA GATGTTCTAA TTTGATGATG CATTTCATGT AAGCGACCTT	6900
10	CTGCAAATGT TTTTGTAATA TAAGGAATTA AACTCACTGC AAAACCAGCA CTTAATGATG	6960
	TCGGAATCAT TACAATTTTA TTAGTTGACA TATTTAGCAT ATTAAAGAAT ATATCTTGTA	7020
	ACTGTGAAGG TATACCAACT AAAGATAAAG CACCGTTATG TGTAATTGA TCTACTAAGT	7080
15	TAAATAATGG ATAATTCAAA CTTACAATAA CGAACGGTAT ACTATAAGCA ATAATTTCTT	7140
	TATACATCTT GCCATATGAC ACATCTATAT CTGTGTAATC AGATTCGACC ATACGATCAA	7200
20	TATTATGCTT ACGCTTTCTC CAGTAATACC AGAGTGTGA TATACCAATA ATCGCACCAA	7260
	CTGCTGCTGC AAAAGTAGCA ATACCATTGG CTAATAAAAT AGAGCCATCA AAGACATTTA	7320
	GTAATAATA ACTTCCGATT AATATGAAAA TCACGCGTGC AATTTGCTCA GTTACTTCTG	7380
25	AACTGCTGT TGGCCCCATA GATTTATAAC CTTGGAATAT CCCTCTCCAT GTCGCTAATA	7440
	CAGGAATAAA GATAACAACC ATACTAATGA TTCTTATAAT CCAAGTAATA TCATCGACTG	7500
	ACCAACCGTT TTTATCATGA ATGTTTCTAG CTAATGTTAA TTCAGAAATA TAAGGTGCTA	7560
30	AGAAATACAG TACCAAGAAA CCTAAAACAC CGGTAATACT CATTACAATA AAATCGATT	7620
	TATAAAATTT CTGACTTACT TTATATGCCC CAATAGCATT ATATTTGCA ACATATTTG	7680
	AAGCTGCTAA TGGTACACCT GCTGTCGCAA CTGCAATTGC AATATTATAT GGTGCATAAG	7740
35	CGTATGTGAA CGGCGCCATA TTTTCTTGTC CACCAATTAA ATAGTTGAAT GGAATGATAA	7800
	AAAGTACGCC CAATACCTTG GTAATTAATA TACTAATGGT AATTAAAAAG GTTCCACGCA	7860
40	CCATTTCTTT ACTTTCACTC ATTACGAATC TCCCTATCTC ATGTTTATTA AAGTTTGTGTA	7920
	AACTAAAAGC TGTTTCTCTG TAAATCATT TTTCAATTAT ATGAATATAT CACAAAACCT	7980
	TATTTCAATTG TCGTATATTC AATGAATTAT CATAACAAAA TTATCAACAC ATTGTCATTG	8040
45	AATACTAGAT TTTGATTAGA ATATTACGAA ATTTTCATATA AACATTATAC TACTATTTGA	8100
	GATGAACATC GCATAACAGT AGAAAAATCA TTCTTATCAT ACACATACAT CTTCAATTTT	8160
	TATGAAGTTC ACATTATAAA TATATTCAAC ATAATTGTCA TCTCATAACA CAAGAGATAT	8220
50	AGCAAAGTTT AAAAAAGTAC TATAAAATAG CAATTGAATG TCCAGTAACA AATTGGAGG	8280
	AAGCGTATAT GTATCAAACA ATTATTATCG GAGGCGGACC TAGCGGCTTA ATGGCGGCAG	8340

55

	GTAAACTCAA AATATCTGGT GCGGGTAGAT GTAACGTAAC TAATCGATTA CCATATGCTG	8460
	AAATTATTAA GAACATTCCCT GGaAATGGGA AATTTTTATA TAGTCCCTTT TCAATTTTTG	8520
5	ATAATGAATC CATCATAGAT TTTTTTGAGT CTAGGGGTGT TAAATTAAAA GAAGAAGATC	8580
	ACGGGCGTAT GTTTCAGTT TCCAACAAAG CACAAGACGT GGTTGATACA TTAGTGACAA	8640
	CTATCGAACG CCAACATGTA ACGATTAAAG AAGAAGAAGC TGTTAGTAGA ATCGAAGTTA	8700
10	ATACAGACCA AACTTTCCT GTACATACTC AAAATAATAG TTATGAAAGC CATTGCTAG	8760
	TGATTGCTAC AGGTGGTACA AGTGTCCCTC AAAGTGGTTC AACTGGTGAT GGTTATAAGT	8820
	TCGCACAAGA TTTAGGTCAT ACCATTACTG AGTTATTCCC GACCGAAGTT CCAATTACAT	8880
15	CAGCTGAACC TTTCATCAA TCCAATCGTC TAAAAGGTTT AAGTTTAAAA GATGTTGAAT	8940
	TGTCAGTACT TAAGAAAAAT GGTAAAAAAC GCATCAGTCA TCAAATGGAT ATGTTATTTA	9000
20	CTCATTTTGG TATCAGTGGT CCAGCTGCAT TAAGATGTAG TCAGTTTGTT TATAAAGAAC	9060
	AAAAAATCA AAAGACACAG CACATTTCTA TGGCAATCGA TGCATTTCCCT GAATTAAACC	9120
	ATGAACAATT AAAACAACAC ATCACATCAT TATTATCGGA CACACCAGAT AAAATCATT	9180
25	AAAAACAGTTT GCATGGTCTA ATTGAAGAGC GCTACTTACT GTTCATGCTG GAACAAGCAG	9240
	GAATCGATGA AAATACCACA TCACATCACT TATCAAATCA ACAATTGAAC GACTTAGTAA	9300
	ATATGTTTAA AGGGTTTGTA TTTAAGGTGA ACGGGACATT ACCTATAGAT AAGGCATTTG	9360
30	TCACAGGTGG TGGTGTGTCA CTTAAAGAAA TTCAACCTAA AACAATGATG TCTAAATTAG	9420
	TTCCGGGATT ATTTTTATGT GGTGAAGTAT TAGATATACA TGGTTATACT GGTGGTTATA	9480
	ATATTACAAG TGCCTCGTA ACAGGACATG TCGCTGGATT ATATGCCGGA CATTACTCAC	9540
35	ATGCATCAAT GGAATAATAG TATAAAATTT GGTTTCGATTC TCTTTAGTAG ATCAACTTTT	9600
	TCATTCAAAT AAAAATGACC TTAATATAAC TGAGTCACTA AAAAGTGTCG TTATATTAAG	9660
40	GTCATTTTCGT TAATTATGAT TCTTTTTTCGT TTTTAGTACG TCTTCTAGCT AACAAAGCCG	9720
	CACCTGTAAT CAGTGCAAAT TCTTTCAATG GTAAATCCAT TCCTTCAGAA CCTGTATTTG	9780
	GAAGTTCTTT TTCAACTTTG CGCGATTCAT GTGTCTCTTC TTTTAAATA GGCCTACAAA	9840
45	CTTTTGAGC TGGCTGAATT TCTTTTGGTG ATACTTTCGT CGCTTCAGCT GGTAATTTAA	9900
	TTGCTAAAT TTCATCAACA ATGAATTGCG TGTGTTGTTT GATGTCATTT AATGTCGCAT	9960
	CTTCATCAAT CATTCTATTG CCATCTGCAA CATATTGATC AATTAATACT TTTACTTTAG	10020
50	CTAATTGTTT TGGTGTGCG ATCGCTTTGA ATTTGCGATA TGTTTGTGTA GCAATGTTAT	10080
	CAATTCGCAG TAAGCTATTT TCTTTTTTCAG TAATTACTGC TTCTATATCG CTTAATGCAA	10140

55

CATCCATTG TAATTTTAAA GCAGTTATAG CTTTAAATGC ATCAGCCTTA TTACGATTAC 10260
 TTAATTTTCG ATAATTTTGC ACTAAAGCAG TGACGCGTGC AAGATCATCA TTAATCGTTT 10320
 5 TTTCAGCATC TGGCTTTTTA ATAGGATGTA CATCTAAATC ATGTATTGTT TGTAGATTTA 10380
 ATGATGCTGT TTTATCAACT TGTGCATTGC TACGATCTTG ATCAATTTGT CCAATAGCAG 10440
 TGTCAATAAT ATTTTGTAAC TGTGCTAATA TACTATTTCT TTCTTCTACC GTTGCTTGAA 10500
 10 TATTCGCTTC AATTGCTTGT TTTTATCGT TGAATAATGT TGTCAATTGT TCTCGAGCAG 10560
 ACGCCTTTCT GTTAATAACA GGTTCGATTT CACGAATTTT GTTTTTCTCA TCATGCAATA 10620
 AATATGCCAC ATCTGCATTA GTCAGTGCAC TAGCAATTTG TTGTTTAGCT TTAATTAACT 10680
 15 CTTTTTCAAC TTGTGCTATT GCAATATTTT GTTCTTCATC TGTCGCTTCG TTATTTGCTT 10740
 TAATTAAATT AATTTTATTT GTAGCGATAT TTTGAATTTG TTGTAATGCT GTTGCTTTAA 10800
 CTGTTGTGCG TGGTTTAATT TTTGAAATAA TATTTTGAGC ATTTATACTA TCTTGATTAA 10860
 20 CTTGGGCAGT CTTATCTGCA TGATTGATCT GATCAATAGC CTGATTAAAGT GCTTGTTCTA 10920
 CTAAATGTTT AGCAGCTAGT CTTTCTTCTT CAGTTGATAA ATCGCTTTGA TCGATTAGTG 10980
 25 CATTTTGAGC TTCGGCTTTT ACACCAACAG ATTGACGCGC TGCTGGTTTA ACTTGAACCT 11040
 TAGGTAAAT CACTTTGATG TTGTCGTTGC CATCAGTCn AGTnCGATCC ACTTCTGCAT 11100
 TCGTTTTGTT TTGTGCAATG TCATTT 11126

30 (2) INFORMATION FOR SEQ ID NO: 172:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3660 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

TTGCCCCGCA CGCGGTGTG nTTCCTAGAA ATAATGAATA TAAAGaGAAA TATATAACAA 60
 CGATTTTGAA TTATGAACCT GGTGATATCG TTACAATCAA ACGTGTGAGA GATAAGACCG 120
 45 ATTTGCTAAT ATATTTGTCT AGTAAAGATA TTTCTATTGG TAATGAAGTG GAAATTGTAT 180
 CGAAAGATGA AATGAATAAA GTAATTATCA TTAAACGTAA TGATAATGTA ATTATTGTCA 240
 GTTACGAAAA TGCAATGAAC ATGTTTGCTG AAAAATAAAA TAAAGAAGCC ATAAAGATAT 300
 50 CCATGATTGA ACTGATAAAG ACATATGGAT AATTGCTTTA GGCTTCTTTT TTATTAGTTA 360
 ATTTATCAAG TGAGTATATT TGAGTAAAT ATTCACTGCA TAAAGATTGA AGATAATCCA 420

55

	CTGTGGACTC GGACGCTGGA AAGTCAATTT AGCAATCGTC CAACTAGATT GTAGAACTTC	540
	GCCTAATAAT ACACCTAAAA TATATTGATA ACTCATTGTG ACAAGTAGTT GAATTTCTAC	600
5	TATATTTTCA TCTTTTAATA TAAAATACAA CATGATAGAA ATTAAAGTTA TAACAACAAT	660
	GGGTGAGCCT TTTCTAGATG TTAAAATTAA AAAATAAATA AATATCAATA AATAGGTAAA	720
	TATAAAGAAA CTAGGTATCT GATAATGGCT CGACGCTAAA CCTATCAATA ACATAATAGG	780
10	TGGCATAAAA TAACCACCAA TCGTTGTAAG CCATTGGCCT GCTAGATGTC TAGATTGTGT	840
	AATTGCGAAT CCTTGTGTGA ATGTCTGTTG TCGCTCTCGT GGACTTGTTA CAATGACTAA	900
	ATCTTTTGCA CGGCCACCAG CGAGTTTATT AAACAGTACA TGACCAAATT CATGTGTTAA	960
15	AACAGGGATA TAGTTTAAAA TGACATCTAA ATAGTTCAAA ACAGGCTTAT GTCTATATTG	1020
	ATGAATAGCA ATATAACAAG CTGCAACAAT AACGATAATG TATATATTAA GTTGAATTGT	1080
20	CGTATTAAAA AAGTTTGATA AATAATTCAT TGTTAACCTC ATATAAGATA TTAATTTAAA	1140
	GTTTGCTTAT CACTTATTAT AAATGATATT GGCATCAATA GCGTTAGACT TTAGACTTAC	1200
	CTTAGTTAAA CTAATTTTAA TTTTGA AAAA GGTGAATATG TGTTAAAATA AAGCAAAATC	1260
25	ATTTGATAT AAATAGGATG AATATAAATA CTGTTAATAT TGATTACACT AACATAATAA	1320
	TGAAATAAGA TAGGAGATTC CTGTTATGAC TGTTGAAGAA AGATCCAATA CAGCCAAAGT	1380
	TGACATTTTA GGGGTCGATT TTGATAATAC AACAAATGTTG CAAATGGTTG AAAATATTAA	1440
30	AACCTTTTTT GCAAATCAAT CAACGAATAA TCTTTTTATA GTAACAGCCA ACCCTGAAAT	1500
	AGTGAATTAC GCGACGACAC ATCAAGCGTA TTTAGAGTTA ATAAATCAAG CGAGCTATAT	1560
	TGTTGCTGAT GGGACAGGAG TAGTCAAAGC TTCGCATCGT TTAAAGCAAC CTCTAGCGCA	1620
35	TCGTATACCT GGTATTGAGT TGATGGATGA ATGTTTGAAA ATTGCTCATG TAAATCATCA	1680
	AAAAGTATTT TTGCTAGGGG CAACTAATGA AGTTGTAGAA GCGGCACAAT ATGCATTGCA	1740
40	ACAAAGATAT CCAAACATAT CGTTTGCACA TCATCACGGT TATATTGATT TAGAAGATGA	1800
	GACAGTAGTG AAcGnAnTTA AACTGTTTAA ACCTGATTAC ATATTTGTAG GTATGGGATT	1860
	CCCTAAACAA GAAGAATGGA TTATGACACA TGAAAACCAA TTTGAATCTA CAGTGATGAT	1920
45	GGGCGTAGGT GGTTCTCTTG AAGTATTTGC TGGGGCTAAA AAGAGAGCGC CTTATATCTT	1980
	TAGAAAATTA AACATTGAAT GGATATATAG AGCATTAAATA GATTGGAAAC GTATTGGTAG	2040
	ATTAAAGAGT ATTCCAATAT TTATGTATAA AATAGCCAAA GCaAAAAGAA AAATAAAAAA	2100
50	GGCGAAATAA TCATGATGAC AAAAATAAAA CCGAGGAAAT CCTTAAATGG AGATTCTCGG	2160
	TTTTTTCGGT TTATTTAATA ACGAAGCGGG ACTCATCGAG TTTGTTTCTA AATTCTTTTT	2220

55

	CATCAAGTTC ACCGTAATCT TTTAACTTTC CGCCTTCAAT CCAAGCAATC TTAGTACAAA	2340
	ATTGTCTCAC TTGTCCTAAG TTATGACTAA CGAAAAAGAT GGTTTTGTTF TGCTCTTTAA	2400
5	ACTCGTAAAT TTTATCTAAA CATTTTTGTG CAAAAGTTTG GTCACCTACA GATAAAGCTT	2460
	CGTCAATGAC TAAGATATCT GGATTAAC TGATATTAAT TGAAAAACCA AGTTTTGCAC	2520
	GCATACCACT TGAATACTTT TTAAGTGGTT GATAAATAAA CTCACCAAGT TCACTAAATT	2580
10	CAATAATCTT AGGTGTCATC GCTTTAATTT CTTTTCGCTT AAAGCCCATA CATAACATTT	2640
	TAAATTCGAT ATTTTCAATC CCTGTAAGTT GTCCACTCAA GCCAGCACTA ATTGCGATAA	2700
	CGCTGACTTC ACCATTACGA TCCACTTTGC CAACAGTAGG CGACAAAGAA CCGCCAATGA	2760
15	TATTGCTCAA CGTTGATTG CCGGAACCAT TGATGCCAAC AAGCCCTATG ACGTCACCTT	2820
	CATATGCTTT TAAACTAATG TCATCTAAAG CGAAAAATGT TTTGTTTTTA TGTTTGGGAA	2880
	TGAGCGCATC TTTCATACGT TCTTTATTTG TACGATAAAT ACGATATTCT TTTGTTACAT	2940
20	TTTTAATGTT TACCGAAACG TTCATTTGTA GACCTTCCTT ATTCACATTT ATCTAGATTA	3000
	TAATATACTA CTCAACAGTT GTTAAATTTT AAAACCTGTT GTAAAGTGTA TAGAAGATTT	3060
25	TGTTATTATC AGAGTGGGTG TTTTGACACA AAATGTTAAT CATCAATGAT AACAAATGATA	3120
	TTTAAAACT AAACCTATTT CAACCTACAT GATTGTATAC TATAATGTAT TTGTAATAAA	3180
	CTAATATTTT AAAGAACTAG ACAATAATTT TGATAGCATC CATGTATAGT GATAGTATTT	3240
30	ACAACAATTA TTATAATACT ATTTAGTTAA GTAGAGAAAT AGTTAAACAT TTGAAAGTGT	3300
	GGTTTAATGG AATGTCAGCA ATAGGAACAG TTTTAAAGA ACATGTAAAG AACTTTTATT	3360
	TAATTCAAAG ACTGGCTCAG TTTCAAGTTA AAATTATCAA TCATAGTAAC TATTTAGGTG	3420
35	TGGCTTGGGA ATTAATTAAC CCTGTTATGC AAATTATGGT TTACTGGATG GTTTTTGGAT	3480
	TAGGAATAAG AAGTAATGCA CCAATTCATG GTGTACCTTT TGTTTATTGG TTATTGGTTG	3540
	GTATCAGTAT GTGGTTCTTC ATCAACCAAG GTATTTTAGA AGGTACTAAA GCAATTACAC	3600
40	AAAAGTTTAA TCAAGTATCG AAAATGAAC TCCCGTTATC GATATACCGA CATATATTGT	3660

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13868 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

	ATTAATCACT	TGTTGTGTAG	AGTCTTGTCC	GTTTTGGTTA	TGATTGTTAG	CCATGATATA	120
	CCTCCCTTAC	AACACTCGTG	GACCAGAAGT	TTTCTGATCT	CTCACATTAA	CTTCTAACTT	180
5	ACGTACTGGC	ATTTCTGTGA	AATATTCTAC	ATTCTTTTTA	ATATCCGAAC	GAATTGCTTC	240
	AGTTAAAGAT	TGAACCTGAA	CATTATTTGG	TACGAAAAAG	TCAGTTTAA	TGTCGATATA	300
	AGATTTATTT	TTTTTGTTAT	ATAGTTTCGC	AACTACATTT	GGTTGTCTTA	CTTGATCATA	360
10	TTTTGCAACC	GTATCGAATG	CCGTCTTTTC	AACAGCTTTA	CGAGATACGT	AAACATGACC	420
	ATCATCGAAG	TCTTTGTATA	ATCCAGGTTT	TCGATGCGTA	GGTTTGAAGA	TACTAAATAC	480
	TAATATAAGA	CCTATTAATA	TCAATAGTGC	AGCAAGTGAA	ATAAGTAATG	GTTGGAACCA	540
15	TTCAAATTGA	AGGAAGTAGT	CTTGATATTC	AGTTATACGT	CCATCTTGGA	TATACATGAA	600
	TAACAGGAAC	CCCACGaTTA	CTACTATTAA	TAAGCCAAGG	ATAAAGTTTT	TAAGTCGTTT	660
20	CACCCCTAAC	GACACCTCCT	TAGTTAAAGT	TAATTTAAAA	ACATATTAAA	TATGTACCCA	720
	TCAGTTTTTT	TCTTAAACAT	AATAAATTAA	TAACTTTAAA	TTTATTTTTA	ATATATAAGA	780
	TGAAGTACCA	TTAGTAATA	TATTCCTAG	TTTTTGTAAG	TAAAACCTCA	TTATTAATTA	840
25	ATTyTCGTCA	ATATGTTTTG	AAGAACGATA	TTCTAAAATA	TCTGGGTCAC	GATGTTTAAT	900
	TAAAACCTTA	TTACTATTTT	TCGGTTTCTC	CTCACTCAAA	GATTTTATAA	GCGACCATAT	960
	CATCGCTATA	ATGACCACGG	AAAATGGTAA	CGCAGCAATG	ATTAATAAAT	TTTGAATTGC	1020
30	TTGAGTACCA	CCTGTGTAAA	TCATGATGAT	TGCAAATAAT	GCCATAATGA	TACCCCAACT	1080
	CACTTTGACA	AATGACTTCG	GATTAATATC	ACCACTTGAA	CTCAACATAC	CTAAAACATA	1140
	AGTTGCCGAA	TCCGCTGATG	TAACAAAGAA	AATCATAATA	ACAAGTAAAG	TAATTAAGCT	1200
35	TAATACAAAA	CCTAGCGGAT	AATGTTGTAG	CGTCGCAAAA	GTTGCTGTTT	CTGTGCGAGC	1260
	TTTAGCAATA	TCGGCAATAT	GATTATCTTG	TAAGTAAATT	GCTGACGCGC	CGAATACCGC	1320
	AAAGAATATA	AAGCAAACCTA	ACGCCGGGAC	AAAAAGTACA	CCTAGAATAA	ATTCTTTAAT	1380
40	CGTACGTCCT	TTTGACACAC	GTGCAATAAA	TATACCTACA	AATGGTGCCC	AAGATATCCA	1440
	CCATGCCCAG	TAAAAGATTG	TCCAATTTTG	TAACCATTGG	AATTTTTGAC	CACCTGTGCG	1500
45	AATGCGTAAA	CTCATACTAA	AGAAATTTGC	AATATAATTA	CCTAGACCAT	TCGTAAATGT	1560
	ATTTAAAATG	TATAGCGTTG	GCCCAACAAT	AAAAAGACCA	ATAAGTACTA	CAAAAGCAAG	1620
	TAACATGTTG	ATATTACTCA	ACGTTTTGAT	ACCTTTATCG	ATACCTGACC	ATGCTGACCA	1680
50	AGTAAATAAT	ATGGTTGCAA	TGACAATCAA	GATTACTTGC	ATCGTGAAGT	TACTCGGTAC	1740
	ATTAAATAAA	AAATGTAAAC	CTTCGTTTAT	TTGCAATGCA	CCGAAACCTA	ATGTTGCAGC	1800

55

	CATTGCCTTT TCACCTAATA AAGGCGTCAA TGTAGCGCTG ACTAAGCCAG GATATCCTTT	1920
	ATGAAAGCTA AAATATGCAA ACACTAGCGC GACAATACCA TAGACTGCCC ATGCATGAAT	1980
5	CCCCCAATGG AAAAAATGAAA ACTGCATTGC ATCATTAAAT GCAGATTGCG TGCCAGCTTT	2040
	ATGAATAGGC GTTAATTTGA AGGCATGACT GATTGGTTCT GCCGTTGTCC AGAACACAAG	2100
	TCCTATTCCC ATACCAGCAC TAAATAACAT AGCAAACCAA GACGGCAATG AGAATTCAGG	2160
10	ATCTTCGCCT TCTTCACCTA ATGTAATGTT TGCCTATCTC GAAAATAAAA TATACACACA	2220
	GACAAATAAA ATAATAAAA CGAGCAATAA ATAATACCAA GAAAAATGTA GCGCAATAAA	2280
15	TGTAGTAATG TTTTGCCTGA GTTTTTCTAA CTGTTTCGGA AATATTGCTC CAAAAGCAAC	2340
	AAATATCGTA CATATCACTA AAGATACCCA AAACACTAGA CTTACTGATT TATTTTTTCAT	2400
	AAATACAAAC CCTTCTGTG TAATGGTAAG TTCATACCCA TAACTGCAAC ATTTTAATCA	2460
20	TTTGTAATTT TATATAGACA CAATTAATAA TGCCTCATCT TTTAAAAATG ATATATAAAA	2520
	CACACTCAAA TTATTTATCA TTGAGCAACA AAGTATTTTA TTGTATTTAA GTAATGCCTT	2580
	TCTAGTGCAT TATTGATTTG ATACCTGCAA AGTTGCCATA TTTCCGTTTA GAATCAATAG	2640
25	TCGCTAGACA CAAAAATAA GTCGCCTATA CAGTATTTTC TGCATAAGGC GACTTTACTT	2700
	ACTAATCTAT ATATTAATTA CTAATTTTCC AATCATTGAT TGTTTTTCCA ACAATTGATG	2760
	TGCTTGATAT AAGTTTTTCAG GTGATAAACC TTCAAAAAC TGTGTCGTTG TTGGTTGGTA	2820
30	ATGCCCTGAT TCTATATTTT TCGTAATATC TTCTAAATAC TCATGTTGTT TAATCATATC	2880
	AGGCGTTCGA TGAATTGGAC GCGCAAACAT AAATTCATGT GTAAATGTTA TACTTTTTAA	2940
	TTTTAATGCA TTTAAATCTT GATCTTCATT AAAAGCTACG ATAGTCGTAA TATGCCCTAA	3000
35	TGGTTTTATC AGTTCAATCA TAGTATTGTA ATACAAGTCT GTATTATAGG TGCAAAATAT	3060
	ATAATCTACT AATGGAATTT CTTTAAATTG ACGCACTAAA TCCTCTTTAT GATTCAATAC	3120
40	GATATCTGCG CCCATCTTTT CACACCACTC TGTTGTTTCT TGTCGTGATG CTGTTGTAAT	3180
	GACAGTTAAA CCATACCGTT TAGCAATTTG AGTGGCTATA CTGCCTACAC CACCGGCACC	3240
	ATTAATGATT AAGACAGACT TCCCTTCGTT TTCAGCAGGA TTCGTAGAAA TTTTAAATGT	3300
45	ATCAAAAAAC GTTTCATATG CCGTAATACC AGTTAGCGGT AGACTAACCG CTTCAATAGC	3360
	ACTTATGTTG TGTGGTGCTT TTGCAACTAT AGCTTCTGAC ACCAATTGAT ATGTCGCATT	3420
	TGATCCTTGT CTATTTGGCG ATCCAGCATA AAATACAACG TCACCCGGAC TAAATAATGT	3480
50	AACGTCTGGT CCGATAGCTT CAACAGTACC AATAGCATCA AACCCAAGTA CACGAGGTGC	3540
	TTGAGTGACT TCCATTTGTC GTTGCTTTGT ATCTACAGGA TTTACACTAA TGCTATTTAC	3600

55

	ATTTCCTTCT	TCCAATTTAA	AGGGCTTCTC	AAATCCTATC	ATTTTCATAT	CGTTTCACCT	3720
	CATTTATGAA	CTTATTTCTT	ATTATACAAA	ATAGAAGCCA	TGTGTGCTTA	TATCGCAGCA	3780
5	TCATGACTCC	TTTTTCATTT	GAATATATAA	ATAATTACAG	ACGACTTTTCG	TATTAAATTT	3840
	TAGACTTATT	TCTACCATGT	TGCTGAACAA	ATTTACTTTA	GATAAAAAAT	TATTAAATTT	3900
	TGGTCAATTA	ACAAAGTTAG	TTTGTTAAAA	CGTgATACTT	TATTATTCCG	TTACTTTAAT	3960
10	AACTTGTTTA	CCAAAGTTAT	CGCCAGTaaa	TAAATTTTTA	AATGCATGTG	GCGCATTTTC	4020
	AAAACCATCT	TCAATGGTTA	CTTGTGACTG	AATTTTACCT	TCTTGAACCC	ATGTTGCAAG	4080
	CTGTTCACCTA	GCTTCTTTAA	AAGCATTAGC	GAATTCACCT	ACCAAGAAGC	CTCTCATCAT	4140
15	TACTTGCTTC	TTAATAAGCG	TACCTTGAAT	ACGTGGTCCG	ATATCGGCTT	CAGGATGATT	4200
	ATATGACGAA	ATTGCGCCAC	ATACTGGTAC	ACGTGCAAAA	CGATTTAAAT	GCTTAAATAC	4260
20	TTCATCGCCA	ACTGTTCCAC	CAACATTTTC	AAAATAAACA	TCAATACCAT	CTGGTACTGC	4320
	TTGTGCTAAC	GCTTCTGCAA	AATCCTCTTT	CTTATAATCA	ATACCAGCGT	CAAAGCCCAG	4380
	TGTCTCTGTT	AAATAATTTA	CTTTTTTGTC	GCCACCCGCA	ATACCTACTA	CACGGCAACC	4440
25	TTTAATCTTA	GCAATTTGAC	CTACAACCTGA	ACCTACAGCA	CCAGATGCAG	CTGAAACCAC	4500
	AACAGTATCA	CCGGCTTTAG	GTTGTCCAAT	ATCAAGCAGA	CCATGATATG	CTGTTTGTCC	4560
	TGGCATTCCCT	AAAACACTTA	AATATAAATC	AAGTGGTACA	TCTGTCGTTG	GAACTTTAGT	4620
30	AATTTGATCC	GCTTGGACAT	GATTAATGAT	TCGCCAAGGC	AACATACCTA	CAACGACATC	4680
	TCCTTTTTTA	TAATCTGCGA	GTGTCGAATC	AATTACTTTT	GCAACGACAT	GGCTAACAAT	4740
	CGGTTTACCA	ATTTCAAAAG	GCTGTACATA	CGAATCTGCC	TTAGTCATAC	GTCCTCTCAT	4800
35	ATATGGATCC	ACTGAAATAT	ACAGCGTTTG	TACAAGTACA	CCATCGCTCT	CAAGTTTaGG	4860
	CGTGfCAATC	TCCTCaATTT	TGAATGTATC	CTCTTGAGGC	ATGCCkTCTG	GTATTTTGTT	4920
40	AAAAAGAATT	TGTTTATTTT	GCATCATTAA	TCACCTTTCT	TTATTTGAAA	CTTTTACTTA	4980
	TTTGTTACTT	AAGCGTTAAG	TTTGAATTGT	GTctTCGTGA	TGTCTGTATG	CAAATACATT	5040
	CTTAGTTGTT	ATATTTTGAC	TTAAGCACTG	ATTCATTCAT	GTAACCTCAA	CCACATTATA	5100
45	TTTGCTATAA	TCATAAATTT	AAAATGTTAC	GACTTAGACA	TTTTATGGAA	ACTCTCAAAC	5160
	AATAGATAAT	TTTTGAAAAG	CTCTAATATT	ACAAGCTTTT	TTGCCCCAGA	AAAACCTAGCA	5220
	GTTGCTTTAT	TTTTTCCATA	AGAAGTCGAT	TAACTCATTA	GCAACATTTT	CATTCTCATG	5280
50	AAGCTGACTA	TGTTGTGCAG	GCTCACCTTC	ATATTTAGAT	TCTCGATAAC	TTTTCGGACT	5340
	ATTTCCAGT	AAATATTTTA	ATGATTTCGA	AGAACTATTA	GACACTCTGC	CGTCTGAATG	5400

55

	ATCTTTTAGC ACGCGTAATT GCTGATAAGG TTGATTCATT CGACTTGGTT TACCATCTTT	5520
	ATCAACTGTA ATTCATTGA CATCTTCATT CATATTTAAA ACACCATTAA ATGTCCCTGC	5580
5	AATATTCACT TGTGTGTTA ACTGTGGCAG TGACTTGTGC TTACCATATG TCATCATATA	5640
	TTGTGCAAAAT GTTAAGTTCC CCATTGAGTG ACCGACAAAG TTGAATTTAT CGAAATTGTA	5700
	TTCAGATTGT AACTTAGTCA GTACATTTTT AAACCACGCA GCATTCTTAT CCAAATAGCC	5760
10	TTGTCTGTTA TTTTCAAGTT CAATTTTCAC AATAGGATTC ACTGCATCTT TTCTTAGTTT	5820
	CCCTTTAAAT GTCACGTCAC CATCCTTTGA AACGTAAGCA GTGATGATAT CTTTAGTTAC	5880
	CCCTCTTTTT TCTGCTTGCT TCACCATAAA CTTTTCAGAA TTGGCACTAC CACCAAATCC	5940
15	ATGTAAGAAC AATGTTGGAA TTGGCTTTTT AACAAATGTC TGTGTGTGTA TTTTAAATGT	6000
	TTGTGCCTGT CGTTGACTAA ACACCACCAT AATAATAGAG CCTATAATAA TAGCGACCGC	6060
20	TAACAATGTC GTAATAATTA CAAAAATTTT CTTCACACTT TTAACCCCA TTCATGTCTT	6120
	TTATATAAGT ATAAAGGATG TGATTAAAAA TGTCTTTTAG TTGATTTTGA ATACATCATT	6180
	AACTTTTAAG ATGACTTTGG AAAGTTGTCC GTTAACGTTT GTTAATTGAT TGCTTCTTTA	6240
25	GCCTTCAATG GTGTGTCACC CATTGATTAA TATATAATA TGTATATGCA TGTTTAATTT	6300
	ATCTCTTTCT ATAAATAAAG ACCTACCAGC ACTCGACTGA TAGGCCTTTT AATATCTATA	6360
	ATTATTTAAT TTCTTTTGTT TCGGCTAACT CTTTGTACCA ATAAGCACTT TTCTTAGGAT	6420
30	AACGTTCTTG AGTCTCAAAA TCGACATAGA ATAAACCATA TCGTTTTTCA TAACCATTG	6480
	ACCAAGAGAA CACATCCATT AATGACCAAA TAAAGTAACC TTTAACATTT GCACCATCTA	6540
	TAATAGCATC TGCAATAACG TTCAAATGTT GTCTTACATA ATCAATACGT GCATCATCAT	6600
35	GAAGTGTGTT TTCAGATTCA ATAAATTCAT CTTTATATCC TAAACCATTT TCAGTGATAT	6660
	AAATCTTATG aTAGTTAGGA TAATCTTTAA CAACACGCAT GaTTTGATCA TATAAACCTT	6720
	GAGGATAGAT CATCCAGTCC CAGTCTGTGC GAGGTACGTC GACATCAAAT TCACGTTGTC	6780
40	CGACACCTTT AAGTTGGTAT TTAGAACCGC CTTTATCACC TGTCGCATTA TGCGTGATTT	6840
	CAGATTCTCC ATCGTAACCT CTCATCCAAT CACTCATGTA GTAATTGATA CCTAAGAAGT	6900
45	CGTTTAAATC TTTGGCTGCA TCTAAAATGG CATAATCTTC ATCTGTAATG TTTAATTTAC	6960
	CGCCATTAAAC AGATAAGATA TGTTGCACAC CTTCCATCGT TTCACGAGAA TACTTACCTA	7020
	AATATGTTGC ATCTAAGATG AATTTATTAT GGATGATATC TTCTAATTCT GCTGCACGAA	7080
50	CATCTTCAGG ATTTGATGGA TCGAACGGAT ATTTTGTGTT CAATGCGTGT ACAACACCAA	7140
	TTTCTCCTTT GTATCGCCA TCTTTAAATA ATTTTACTGC TCTAGCATGA GCCACCATCA	7200

55

	CTACTAAATA TTGACCATCA CCAATAGGTC CAATTTTCATT GAATGTAGTC CAATATTTTA	7320
	CTTCTGGGAA TTCTTTAAAA CAATATTCAG CATAATCTAC AAAGTAGTCA ATCGTTTTAC	7380
5	GATTTAGAAA ATCGCCATCT TTGTGTAAaCA CTTCTGGTGT ATCAAAATGA TGCAATGTTA	7440
	CAAATGGTTC AACATGACGT TTATGACACT CTGCAAATAA CTTATGGTAA TACTCAACAC	7500
	CTTTAGGGTT AACTTCGCCA TATCCATTTG GGAAGATACG AGACCATGCA ATTGAAATTC	7560
10	GGATACCATT AACACCGAAT TTTTCACTTA ATTCTAAATC CACTGGATAT CTGTTATAAA	7620
	AATCACTCGC TGGTTCTGCA GTGTACCAAT AGTTTTCTTC TAAATACGTA TCCCATGCTA	7680
	CGCGACCTTT ACCATCCGTA TTTGTGCGAC CTTCTGCTTG ATATGCTGCT GTTGCTCCAC	7740
15	CAAAAATAAA ATCTTCAGGT AATGTTTTAG TCATATGAAA AACTCCTATT CTTAATTTTC	7800
	AAATGTGTGT TGAACGAAAT CAAGGGCTGC TTGGCCATCT CGTGTCAATT TGATATATTC	7860
20	AGCACCTTGA GTCTTCGCTA ATTTAATACC TAATCTATCT GTATCTTGCT TAATATCTTC	7920
	ATAGTTAGAC GCAACTTG TG GCGCTAAAT GATTAATTGG TACTCTTTCA TAATGTCCAT	7980
	ATGTGCGCCA TATCCGCCAG cTGCCGCTTT CACTGGCACA TGATATTCTT CAGCTGCTTT	8040
25	ATTAAGTGCA TTGGCTAATA ATCCACTTGT ACCACCACCG GCACAAAGTA CTAAGACATT	8100
	TGTTTGTCT GTGATATTTG AAGCTTTAGC TGCATCGTCT GATACACCAC TTGCCGCTAA	8160
	AATTGAATCA GCTTTTTTCG TATCAAAGTT TGCTGCAACT TTTTCTTTTA AATCTGAATT	8220
30	ACTTTCTTTA CGTCCTTCTT CTTTCATCAAG AATTTCACTA TCATAAACTT TTAGGAATGG	8280
	GTAGTAAATA ATAATATCTA CAACAATCAA AGTAATAGCT AGTACGAATG ACCATAAACC	8340
	AAAACCTGTA CCCATGATAA TGCCCAATGG ACCTGGTGTG GTCCAAGGTA AATTCACACT	8400
35	AAAACCTGTA CCCATGATAA TGCCCAATGG ACCTGGTGTG GTCCAAGGTA AATTCACACT	8460
	TGCTAATACA AATGGAATAA AGAACACAGG ATTCAATACT AGTGGTGCAC CAAATAAAAT	8520
	TGGTTCGTTT ACACCAAAGA ATGTTGGTAC AACTGATGCA CGTCCAATCG CTTTGTTCG	8580
40	TTTAGATTTT GTCATCCACA TAAACATGAA CGGGACGACC AATGTTGCAC CCGTACCTCC	8640
	AAATGTAACG ATAAACATTT GTGTACCTGA TGTAATAATT TTATCTGCGT GTTCTCCAGC	8700
45	TTGAAGCAAC TTGAAGTTTCG CTTTCGATATT CGCATATGTA ATGGCTGCAA TTGCTGGCTC	8760
	TACAATTGAC GGACCATGAA TACCTACAAA CCAGAATAAT GCAAAGGCAC CAAAGATAAT	8820
	TGTGACACCA ATCCATCCAT CTGCTGCTGT AAATAATGGT TCGAATAATT TTAATAATACC	8880
50	TTCCGCTACA TTTGATTTAA AGCTGTTGCG AATGACTAAA TCTAATGCAT AAAGAATGAT	8940
	GATTACCGCT GAAAATGGAA TTAAGTCCTT AAATACTTGT GAAATATTCG GCGGTACTTC	9000

55

	AAATGCTGAT AAGAATGCTG TTAATAAACC TTTAGTTCCC ATAAATGCAC TTAAGAATCC	9120
	ACCATCTTTG GCTGGATCAG AAGCTAAGAA CAAGAATCCA CACATCGCTG CTAGCATTGT	9180
5	AGAAATAAAG TTAATTTGAT TTGTACTTTC TAGCTTACGG TTAAATGAAT CTGTTAAAGA	9240
	TTTCGCTGTC GTTCCTGCTA CTAAAAATGC TACAAGCCCC ATCGTATAGT TATATGGTTT	9300
	CATTAAAAATG GCTTCCATGC CTTTATCCCA TTTAAAACCA AATATATTTG GCACATATGC	9360
10	AATTAATAGA AAGATACTTG AGAATAAGAT GACAGGCATT GCAGAAATAA ATCCATCACG	9420
	GATGGCTCTT AAATATATGT TACGTGATAA TTTCTCGAAA AATGGCTTCC CTTTTTCAAT	9480
	TTGTGCGATC AATTTTTGCA TCATTGTCAT CACCCTCTTT TATAAAATTC TAATAAATGC	9540
15	TTCAATAAAT CTTTCAGTAA AATTGTTGTC ATTAAATGAT CTTGACCATG CATCATCGTT	9600
	ACACTGTATG CAATATCATC ACCTTGCGCT TCTTTAGCCA ACAGGCTTGT TTGTGCTCTA	9660
20	TGCGCTTCCG CAATGCAATT GTTTCCTTCT TCAATCAGTG CATCTGCTTT TGCAAAATCT	9720
	CCAGCTTGAG CTGCTGTAA TGCTTCTAAA AACTTAGAAC GTGCATCCCC TGCAAAATGCA	9780
	ACAATTTCAA AACCTAATAA TTGGACTTCT TCTCTATTCA TAGCATTAAAT CCCCTTTTAA	9840
25	ACTTATTTTC TTTGTTTCCA AGATGTCGCA GTATCTTTTA ATACTTTATT TAAGTCATCA	9900
	ATATTTTGA AACCAGTTGT ACGTAACCAT TCACGAGCAG CATCTTCACC TTGTTCAATG	9960
	TATACTTGAA CAGCACCAGA CCATGTAGCA CGGCCACAAA GTACCCCGTT GAATTTAGCA	10020
30	CCAGCTTCGT GAGCAAATTT TAAAGTTTCT TGGAAATAATT CCGCAGAAAC ACCAGCACTT	10080
	AAGTAAATGT ATGGTAAATG AGTTGCTGCA TCTTGATCTT TAAAGTGTTG TGCCGCTTCC	10140
	TCTTTTGTAT AAACCACTTC ACCTTCAGCA AAGCCTTCTA CATATTTTCA TTTTACTGGT	10200
35	ACTTCAACTT TCAATACATC AACGTAAAG CGTGGTCTG AGAATAAATT CATTGCTTCG	10260
	TTAACTTTTC TAGGCTTAAC TTTTGCGAAT TcAACAcTAC CGTTATcAGG AATGTTGTCA	10320
	TCGTATGTTA ATACTTCTAA AAAGAATGGA ATATCTTCTG CAACACATTc TGAACCGATT	10380
40	CTTTCATGT ATGCTTTCTT TTGAATGTTA ATTTCTTCAG CATCATCAAC ATCATAGTAA	10440
	AGTAAGAATT TAACAGCATT TGCGCCTTGT TCTTTTAAAC GTTTTGCAGA CCACTCTACT	10500
45	AAACAGTCAG GTAAACGACC TTTAGCGTTT ACGTCATATC CAGTTTTTTC ATAAGCAAGT	10560
	AATAATCCAC AATCTTTGTT ACGTGCATCT GAAGCTGGTA AACCATATTc AGGATCTAAT	10620
	AAAATGAAG ATGCATATTG TGTTAATTCT TCCGCAACTA ATACTTTTAA TTGTTCAATT	10680
50	TGAGCTACAG TTGGTTCTTC AGTTTGATGT TTTGCCATCA TGCGTTTTAA AGCACCACGT	10740
	TGGTCAAATG CTAATGCAGA AATGATACCT TCGTTGTTAC TTAATTGTTc AATTGATGCG	10800

55

	TCATCATAAT TATTTAAATT GACATAACCT GTTTGTGCTT CTTGTGCATT CAGCATGCCT	10920
	AAAGTATTGG CTTTTTTTAG TAAATCGTGG TCGTTTTCAT GATTAAGAAT TGCTGAAGTA	10980
5	ATTCCAGCAA CTGTAGAATC ACCTGAACCA ACCGGATTTA ATACACTTAT TGTCGGAATA	11040
	TTCACCTCTAT AGAATGTATG ATTGTGCTTA GCGAATGCAC CTTGTGCACC TAAAGACACA	11100
	ATAATCCACT CAATCCCTTC GAATAAGGGT TGTGACACTG CCTGTTTCAA ACTTTCTAAA	11160
10	CTTTCATCAA GTGGCTGGTT AAGCAATTGA TATAGTTCAG AAATGTTTGG TTTAATGACT	11220
	GTAGGTTTGT ATGGATTTTC CAAAAGTGTG TGCAAAGTtG CACCCGAGCA ATCTAATATC	11280
	ACAGGCACAC CTTTGTTTTG GCATCGTTCA ATGATTTGTG CATAATAATC TTGATTTAAT	11340
15	CCTTTAGGTA AGCTACCTGA AATAGCAACT GCTTCAACTT TTTCTAATAA TTGTTCAAAA	11400
	TGTTTAATAA ATCCTGCAGC CTCTTGATTA TCAATCTCCG GTCCCTGCTC TAAAATTTCT	11460
20	GTTTGTGTC CTTTCATGTAA AATTGCAATG CAGTTTCGTG TTTACCCTT AATGTTATAA	11520
	AATGCATGCT TGATGTCGGC ATGATCTAAT TTTTAGCAA TAAATTGACC TAATTCACCG	11580
	CCAATAAAAC CACTCGCAAG GACTGGCTCA CCTACTTGCG CAAGTACTCT TGTTACATTT	11640
25	AAACCTTTAC CACCAGCTGT TTTACTTACT TCTTGAACAC GATTAACATC ATCTAATTTT	11700
	AATGCTGTTA ATGGGTATGA AATATCAACG GATGGATTTA ATGTTAAAGT TAAAATCATA	11760
	TGTGTCGTCC CTTAATCGTG GTATTCGCCT CTGTCCCATT TTTCTAAGAA TTCATCAAAG	11820
30	AAATGTGGAT CAGCTTGATC TGCATTGCTT GTTTCTAAAT GTTTAATTTT AGCGATTAAT	11880
	TTTTTGTTCT CTTCAAGTTG TTTATATTCA GCATTAATAA ATGCATCGAT AATATCGCAC	11940
	ATTAATAACT CACCTATAAT ACGTCCACCG AAGCCAATAA CGTTCGCATT TAATCTTCT	12000
35	TTAGCGTATA ACGCTGATGT CATATCACGT ACTAGTGCTG AACGAACGCC AGGTACTTTA	12060
	TTTACAGCAT TGTTAATACC AACACCTGTT CCACAAATAC AAACACCTAA GTCTGCATTA	12120
40	CCGCTAACAA CTGTTCGCC AACTTTTTTA CCAAAAATTG GATAATGTGT TCTTGTGAAA	12180
	TCGTATGTTT CTACGTCAAT GACTTCATGT CCTTTTGATT TTAAAAATTC AGATACACGC	12240
	ATTTTTGTAT CTGTAACAAT ATGGTCGCAT CCTAATGCAA TCTTCATAGT AATTTTTCCT	12300
45	CCTTAGCACA TTTTATTAAG CATATCTACG CGGATTTGGT GTCTACCACC ATCGTATTTA	12360
	CCTTCAACAA AACCTTTAAC GACATTTTTC GCTAATGTGT CTCCAACAAT TTCAGATCCC	12420
	ATAGTGATCA TTCTTGAATT GTTATGGCCT CTAGTCATAT ATCCAGAGCG TTCATCTGAT	12480
50	ACTTCAGCAG CAATCATGCC TTTGATTTTT GTAGCAACCA TAAAGCTACC TGCACCAAAT	12540
	GCATCGATAA CAATACCTAA GTTACCTTCT TGACTTTGAA CATCTTTTGC TACAGCCAAA	12600

55

TCTAATAAGT ATGATTTGAT GACTTCTTTT AATCGTTTGC CAGCTTCATC TGAACCAATA 12720
 ATAATCGCCA TAATAAGACT CCTTTTACT TTAATTTTGA AATACCTTTC TTAAATGTG 12780
 5 ACATATTTAT TTGTAGGTTA TGAAATCTT GAGAAAAGGC TTTCAATTTG ATTACGTTTA 12840
 AATTATAAAC ATAAACAAAC AATAAATCAA CATAATATGT TTATAATATG TTTGTTTATG 12900
 ACGTATTTTC AAACAATAAG TGAACATTCA TATTGTGGTG TTGTTTAAAT TAGGTATTCG 12960
 10 TCTGAAATTG TAGTAAACT TTGTCGAGGT TCCCGTTGac ATAAATTTGC ATAAAAAAtA 13020
 GCCCATAAAT GAATGCAAAT TCACATTAC TTATGAGCAT ATAGATACAT ATTTTAACAA 13080
 TGCAGTTATA CTTTAAATTT AGTCGACTAC TTCAATATAT GTTTTAAATCG TTTCTACTTT 13140
 15 TTCTTCATCT TCATAGTCCA TGACCACTGC AGTCAATTCG TTTAACTGAC AAAATGATGT 13200
 AAAATCTTCT TTGCCAACTT TCGTATGATC GATTAACAAG TATTTTTCAA TTGAATTACT 13260
 TAGTGCCAGT TGTTGCGTAT AGGCTTCATC TAATGTAGAT GTCATCACAG CACCTTTATT 13320
 TACTGCGTTA CTACTAAAGA ACATCTTGCT AAATCTTAGT TTTTCCAACA TGGCGTTCGC 13380
 cATTTACCT ACAATGCTT CTGTAATATG GCGCATTTCA CCACCAATTA AATAGACACG 13440
 20 AAAATGTGCT GTTTGTTTTT CTAACAAAAT TTTATACACC GGCAACAAT TCGTAATAAT 13500
 TGTGAGCGTA TGATGATTGA CTTCTTCTGC TAATAGTTCC ACTGTTGTTT CTGGTCCGAA 13560
 AAACAAAGTA TCCCCATCTT CAATTAATGA TGCAGCTTTT TTAGCTATAA ATCGTTTTTC 13620
 30 TGCAATTTGA CGGTATGTT TTTCTTTATG CGATATTTCT TTATACTGAA ATGTTGAATT 13680
 ACTGCGTGCA CCACCATGAA TCTTCGTAA AATCCCTTTA TTTTCCAATT CAATTAAATC 13740
 TCTTCGAACT GTCATATCAG ACACATTTAA ACCTTCGACG ATTTCAATTCG TTCTTATCGT 13800
 35 GCCCTTTTTA TTCACTAGTT TAGCAATTTT GTCCAAACGT TCATGTTTAT TCAATGTAAA 13860
 ATTGCTC 13868

(2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4549 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

50 TTAAGTCAAC TTTGTCTATA CGGTTTGGAT CtTCTaCCCA ATGTCTTATA AAAGACAATC 60
 CCGCACCTGA AACATAACTC ATGAAATAAG AAAATGGTAT ACCATTAAAT TGATCATTTT 120

	AATCTTTACC	CATACGAAAC	ATCAATTGAT	AAAATGCGAT	GTCTTTTTCT	ATCATTTCTA	240
	TTAAAACGGT	CATAATTGA	TGTATGTTAT	CCGTGGATAA	CTTAACTGCT	CCATTTAACT	300
5	TCTCATCATG	AATGAAGTCT	CTTATTTCCCT	CCAACTGCTG	GTCCTCTAAT	TTTTCAAGCA	360
	AATCATACTT	ATCATAATAA	TGCGTATAAA	ATGTACTACG	GTTAACATCA	GCTAAATCTG	420
	CAATTTGTTG	CACAGTAATC	TCTTCTAATT	GGTGTGTATG	TAAAAGTTCA	ATAAATGCAT	480
10	TTCTCATTGC	AACTTGTGAT	TTTCTAATAC	GTCGATCTAT	AGTCATTTAT	ATCAAGTCCT	540
	CCCCAATGAT	TATAAACGTT	ATGTTCAATTA	TCCCACAAAT	CTCCAACATT	GATGATTGGC	600
	ACACAATGTT	TACCTGTTTA	ATATAGGTGA	TACAAACAAA	CAGAAAAAGG	TGATAACAAT	660
15	GAACCAACAT	TTACTAGGAA	ATCCAAAATT	AACTGTAACT	CATGTCAATG	AAGTTAAAGC	720
	CGGTATTAAAC	CACATCGTTG	TCGACAGTGT	TCAATATGGA	AATCAAGAAA	TGATTATGGA	780
20	AAAAGATGTC	ACTGTGGAAA	TGCGCGATGG	CGAAAAATTA	TATATTAATA	TTTTCAGACC	840
	AAATAAAGAT	GGCAAATTCC	CTGTAGTTAT	GTCTGCAGAT	ACTTACGGTA	AAGATAATAA	900
	GCCTAAAATC	ACAAATATGG	GTGCCCTTTG	GCCAACATT	GGTACCATT	CGACATCTAG	960
25	TTTTACACCT	GAAGAATCAC	CAGACCCAGG	ATTTTGGGTG	CCAAATGATT	ATGTTGTAGT	1020
	TAAAGTTGCA	TTACGCGGTA	GTGACAAATC	CAAAGGCGTC	TTATCTCCAT	GGTCAAAAAG	1080
	AGAAGCGGAA	GATTATTACG	ArTGATTGAA	TGGGCAGCAA	ATCAGTCATG	GAGTAATGGA	1140
30	AATATCGGGA	CAAATGGTGT	TTCTTATCTT	GCGGTGACTC	AATGGTGGGT	CGCATCATTA	1200
	AATCCACCAC	ATTTAAAAGC	AAtGATTCCCT	TGGGAAGGCT	TAAATGATAT	GTATAgAGAA	1260
	GTAGCCTTTC	ACGGAGGTAT	mCCAGATACT	GGCTTTTATC	GTTTCTGGAC	TCAAGGTATT	1320
35	TTTGCGAGAT	GGACAGATAA	TCCAAATATC	GAAGATTTGA	TTCAAGCACA	ACAAGAACAT	1380
	CCTCfGTTTCG	ATGATTTTTG	GAAACAGCGT	CAAGTGCCAT	TATCACAAAT	TAAAACACCT	1440
40	CTACTAACAT	GTGCTAGTTG	GTCTACACAA	GGTTTGCACA	ACCGTGCTC	TTTTGAAGGA	1500
	TTTAAACAAG	CTGCATCTGA	AGAAAAATGG	CTATATGTGC	ATGGACGTAA	AGAGTGGGAA	1560
	AGTTACTACG	CTAGAGAAAA	TCTCGAACGC	CAAAAATCAT	TCTTTGATTT	TTACCTTAAA	1620
45	GAAGAAAATA	ACGATTGGAA	AGATACGCCT	CATGTCATTT	ATGAAGTTAG	AGATCAATTT	1680
	TATAAAGGCG	AATTCAAATC	AGCGTCACGT	GTCCCTTTAC	CTAACGCAGA	ATATACACCA	1740
	TTGTATTTGA	ATGCTGAAAA	TCACACATTG	AATCATGCAA	AGATTAGTAG	CGCGCATGTC	1800
50	GCACAATATG	ACTCTGAAGA	TAAACAACAA	GATGTAAGTT	TTAAATATAC	GTTTGACAAA	1860
	GATACTGAGT	TAGTTGGAAA	CATGAACTTA	AACTATGGG	TAAGCACTAA	AGACTCAGAT	1920

55

	CCTGATTTTA	ATCATATTGA	AAATGGTCAA	GTAGCTACTG	GTTGGTTACG	CGTATCACAT	2040
	CGTGAATTAG	ATCAAGAAAA	ATCCTCAATC	GCGCAACCTT	GGCATAAACA	TGAAACAGAA	2100
5	TTAAAGTTGT	CACAAGATGA	GATTGTACCT	GTTGAAATCG	AATTGTTACC	TTCAGGCACG	2160
	CTATTTAAAC	AAGGCGAAAC	ATTGGAAGTT	GTTGTAAAGG	GTAGTGAAAT	TGTAATTGGT	2220
	AATAGTACTC	CTGGCATGAA	AACACGTTAT	GAACATGAAG	AAACCGTAAA	TAAAGGCATG	2280
10	CACATGATTT	ATACTGGTGG	TAAATATGAT	TCACAAITTA	TCATTCCCTAT	CGTTAATTGA	2340
	TATGCAGCAA	TTACGGTCGC	TTTTGATTAA	AAGTGACATA	GTGATAGGAC	TGTATAACAA	2400
	GAGAAAGCCA	CACGCTTGGA	ATCTTAAACC	AAGGTGTGGC	CCTTTTTATT	ATTGATGGCT	2460
15	ATTTAATTTT	ATAACACTAT	CGTATTTTCT	TTTTCATGAA	TCATTTCAAT	AATGACATT	2520
	TCTTCATTCA	TTACTGCTAC	TTTAGTGCA	TGGTTTTTAA	TTTCTTCTTC	ATTCAACTGT	2580
20	GCATAAGTCA	TGATTATGAC	TACATCGCCT	ACTTCAACAA	GTCTTGACGC	TGCACCGTTT	2640
	AAACAAATTT	TACCACTACC	TCTTTCACCA	GCTATTACGT	ATGTTTCAAA	ACGTGCACCA	2700
	TTATTATTAT	TCACGATGGC	TACTTTTTCA	TTTGCAAGA	TGTCTACCGC	TTCCAATATA	2760
25	TCTGAATCAA	TCGTAATGCT	ACCTACATAA	TTTAAATTTG	ACTCAGTCAC	TCTTGCTCTA	2820
	TGAATTTTAG	CATTCATCAT	TGTTCTTATC	ACTTTATTCA	GCTCCAATTA	TTATATTATC	2880
	TATTAACGC	GCTTTTGAAA	ATTTAACAGC	TAAcGAGATA	AATATGCGTC	CAGTTATTTT	2940
30	GTGTTGTCT	ACTAATGAG	GATACTATA	AACAGCAACT	TCTTCAATGC	GTTCACTTAT	3000
	ATGTGATTCA	AGATAITCAG	TAACCCTGTC	TATAATTACT	TTACTTTGAC	GTTCAACGTC	3060
	TTGATACAAC	GCTTGCTGTA	ATAGCAAAC	TTTACTTAAA	TGTACCGCTT	CTTGTCGTTT	3120
35	TTGCTCCGTT	AAATAAACAT	TTCTTGAAC	TTTCGCCAAA	CCATCTGCTT	CTCGAACGAT	3180
	ATCAATACCA	ATAATTTCAA	CGGCATGATT	GAAGTCTTTT	ACCMTTTGCT	CGaCAATAGC	3240
	CAATTGCTGG	GCATCTTTT	TACCAAAATA	AGCATAATCC	GGCATAACAA	TATTAAATAG	3300
40	CTTATTAACT	ACTGTTACCA	CCCCATCAAA	ATGCCCTGGr	CCGtTCGCTC	CTTCTAACAC	3360
	ATCAGCTAAT	GGGCCTACTT	TGACATCAAT	ACCTAATTCA	CCTGGATACA	TATCTTCTAC	3420
45	TGCAGGATGA	AAAACAATGT	CCGCTCCTAC	TTCTGATACT	AATTCTAAAT	CTTTATCAAT	3480
	TTGTCTCGGA	TAAGCATCGA	AATCTTCGTT	TGGACCAAAT	TGTAATGGAT	TAACAAATAC	3540
	ACTCACAATT	GTAATATCAT	TTGTACTAAC	TGATTGCGGT	ACCATCGTTA	AATGTCCATC	3600
50	ATGTAAGGCA	CCCATTTGTT	GGATAAAACC	AATCGTTGTG	CCTGAGCGTT	TGGCTGCTTT	3660
	AACAATGTGT	TGCATCTCTT	TTACCGTAGT	AATCAGCTTA	GTCATTGTTA	TTAACCTCAT	3720

55

GATCGTATTG TTTTAAACCA TCCACACCAA CACTAAAATC AGCAAATTGC TTCACAAATT 3840
 TCGCTTTTATG TTCAACACCA TAATTTAACA TATCGTGATA AACCAATACT TGACCATCTG 3900
 5 TACCTTTTCC TGCACCAATA CCAATGACTG GAATTGTTAA GTGCTTGCTA ATTTCTTCTG 3960
 CTAAATCATT TGAATTGCT TCAAGTACTA ACGCAACTGC ACCAGCTTGT TCTACATTTT 4020
 TCGCGTCTAA AATAAGTTGc TCCGCTGCTT CTTCGTTGC ACCTTGTAAT TTATACCCCA 4080
 10 TAACGCCAAC ACTTTGAGGT GTTAATCCTA AATGTGCAAC AACAGGAATA CCAATTGCCG 4140
 TTGCTTTTTC AATAAATGGT GTAATATGCG CTCCTTCTGC TTTAATTGCA TTTGCATTCTG 4200
 TCTCCTGATA AAGCTTTAGA GCATGATTTA AGTCTTGTGT CATAGAGATG CCTACTGCAC 4260
 15 CAATCGGCAT ATCAACAACT ACAAATGTAT TTGGTGCGCC TCTTCTTACT GCACGACCGT 4320
 GATGAATCAT ATCTGCTAAC GTCACCTGTA CGGTACTTTC ATAACCTAAT ACAGTCATAC 4380
 20 CAAGTGAATC CCCAACAAGA ATCATATCAA TACCCGCTGC TTCCACTTGT TTAGCACTTG 4440
 GAAAATCATA AGCTGTTACC ATAGAAATTT TAGTTTGCTT TTGTTTCATA TCTATTAATT 4500
 GACTTACTGT TTTCAATGTT ATTCAACCTC TTTTTCAGT ATnATTAGA 4549

25 (2) INFORMATION FOR SEQ ID NO: 175:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8339 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

TTATCTTTTG TTGTTTCCTT AGACAAACGA CTAACCACAT TATAATGGAC TAATTTATTA 60
 ATTFTATTTA ATTCCATTAA GTTATCCGTA AACTAAGTG AAGATGCGGA GTTCACTCTC 120
 40 GTTTGTACTC TTCGTTTAA TAAAGCACCT CGTAATAATA CAATCATTCT TCTTATTAAT 180
 GATGCTTGTC TATATACCTG TGTTCTTTCA GCATAACGCA TATAGTTTTC AAGTACACTA 240
 TTCGTTATTT GTCCTTCATC TACTAAAGAC TCTAATGTTT TkGTTTCTAC ATTAAAAGCA 300
 45 ATTTTITGTA GACGTTCTAA TTCTTTAGAG TTTTCATCAT CTTTCTCTAC AGTTTTTAAA 360
 AATGCTAATT TATCATGATA TTCTTTAATC ACGTTACCAT ATTTAAACT TGTTTCGAAA 420
 GTAGATTTTT GATTTAGATA ATCAATAACT TGTTCTAATA TATAAATCT AGCAACTTTA 480
 50 AACGACATAT TGCCAATTAC TGTTTTAGGT GCAGGTTTCG TTAATAATGG CAATAATACT 540
 TGCGCAACTA CCAAATAAT AATAACCATA CCAGATGCAA TAAATAATAA GTCGTTTCTA 600

55

	ATTGTTCCAT GCACACCACA TAACGTCATA ATTAAAGCGT ATAAACTTCG CTTTGGTGGT	720
	TTCTCAGTCG TTGGATTATC ATCATCATTT TTAGTCATCA TTTTGGAA TGGACTGATG	780
5	GCTAAATAAA AATAAGGATA TAAGACATAA ACCCAAACAA ATCTAAATAG ATAGACAGCT	840
	AAAGCAACAA CAATAGTGAT GCCTATTAAA AAGATTAAAT TGTGCGGTTT TGTTTTGATA	900
	ATTTTAATAA TAACTTCAGG TACTAAAAAT CCTAATATTG AAAAAACAA GCCATTTAAA	960
10	ACATAACCTA GTATATTCCA TGTATGATTG TAACTCATTT GCAGTTGTGT ACGTACTTGC	1020
	ATAATTCTGT CACGTTTCGAA ACCATGTACA AGTCCTGCAA CTAAGTCTGC AATGATTCCT	1080
	GATGCGTGaA ACAATTCAGC AATTAAATAC GTAACAAATG GTGTTAACAA TTGAATAATT	1140
15	GTAACATAT TAATGTTTTT ATATCCTCGA CGCATCAATG TTAATCGGAA CCTTACTAAT	1200
	GCCATACCTA TAAGTAAACC AACCACTGCG CCACCAATTG ATGCAATTAA AAACAACGA	1260
20	ACAGCATCAA CAAGTGAAAA AGCACCTGTA ACTAATACTC CAACAGCTAT TTAAATGAA	1320
	ATAATACCAG CAGCATCATT CAATAATGAC TCACCTTCAA GAATTGTCAT TGCTCCTTTT	1380
	GGCAAGACCT TTCCTTTAGT GATTGCTTGC ACTGCTACTG CATCAGTAGG ACAAAGAATG	1440
25	GCAGCAATTG CAAATGCTGC TCCAATAGGT AAATCTGGCC AAATCCAATG AATAAATAAA	1500
	CCTACACCTA TCACAGTAGT AATGACTAAT CCTAATGCCA TCATCATCAC TGGCTTAATA	1560
	TATTTCTTA AATGGACTCT AGAAACATTA ACACCTTCTA CAAATAACAA AGGCGCAATC	1620
30	ATTGTTACCA TAAACAATTC AGAATCAAAA TTAAATTGAA CAGGGATTGG GGTAATAAAT	1680
	AGTAACATGC CCAAGAAAAT TTGTATAAAT GCTAGGGGTA CTTTAGGTAT GAAAGTATGG	1740
	ACAAACGAAC TTAGTATCAC AACAGCTATA AATATAAGAA TTGTTTCAA TATTTCCAAA	1800
35	CTTTCACCTC TCTAAAAAGT ATTGTTTAAT TGAAAATTAA GTATCACATC TCGTTGTAAT	1860
	TATACTTTAG AGGATAAATT GAGTTAGCGA CCACAAAAGC ACTTTAATAT AGATATATGT	1920
40	CTACGATTGC AGTACTTAAA TTGCAATTA TTTAATTTTA TTTATCACT AATTGTTTGT	1980
	ATAAATAAAC AACTTGCTTT CACATAACAA CATTAACCTA TAATACAAA AATGAGCACC	2040
	TTAAAATCGA CTAACCAATT TCaAAGTACT CTTTAAATGA TTAATTTTGA AAACAGATTT	2100
45	TCaAAAGCAT TGTATGCTT AACAATTTAG CCAACACTTC AATCGTTTTG ATACCATTC	2160
	TTACGATGCT CTTCTCGTTT TTCAGCACGT AATGTAAATG CTTCTGTAGA GTTTTGTTC	2220
	TTTGAACCTA ATAATATTGA TGCATGTGTG TGAGCATCAT TTTTTCGATA CATATAAGCG	2280
50	CCGTTGCGAT AAGCAGCGCG AGCGACTAAG TGCATGCCGA CTGGTGAAGT TAAATTAATA	2340
	AAAACAAGTG ACAGTAATAA ACGCACACTG AAAAACTCTG TATTCACAAT AAAATAAATT	2400

55

	CTTAAGAAAA CATCTTGGAA TTTCACGATA CCTATTGCAC TAATAAGAGC AATAAAACTA	2520
	CCTAACAACA ACATCACAGC AGCAATAAGA CTAAAGATTT CTTTGTAT TTCCATTAAA	2580
5	CACATGCCCC CCACCAATAA AGCGTGATAT TGAAACAGAA CTTACAAAAG ATATAATGGC	2640
	AATGAGCATG ATTGAATCTA AGAAAGAAAC GGTGCCCATA AGTACACTTA ACACACCCAC	2700
10	AATTGACATT ACGACAGCAC TTGTTGTATC AAATGTAACG ACACGATCTG CTGTTGTAGG	2760
	TCCCTTGATT AATCTAAATA AACAGATGAT TAATGCAATT CCAAAAATAA TGAGTGAACT	2820
	AATAATCATA ATATGTGTTA TTGTTTGTAT CATCGCGACA CCTCCAATAT TAAGTCTTCA	2880
15	TAATGCTTAA TACTTCTTAA CAAACTATCT TTTTCTTTT CTGACACGTC GATACTATGA	2940
	ATAAAAAACT TTTTAGAGTC TTGAGAAATT CGTATTACTG TAGACCCTGG AGTTATAATA	3000
	ATTAAAATTG TTAATAATGT TATTGACCAA TCACTTGTTA GTCTTGTTTC ATATGAAAGT	3060
20	AATCCAGGGT TCATATCTTT TGTTTTAAAA AGAATATAAT TAATCGTGCT AATGCTAGAT	3120
	GTTATTAATT GATATAAATA AACACCTAAA AATTTAATAG CTACCCATAT TTTTCTAACA	3180
	TAAAAATCAT CGCTGAAAAA CCTGTGTAAT ATATAAATGA CAATTAAACC AATTAGATAT	3240
25	CCAGAAAAGA AAGTCGAGAA TTTAAATGA TCTTCATCTT GAAATAATAC CCATAAGAAT	3300
	GCAATGATAA TATTTAAAC TATTTGATTC ATTTAGTCCT CTCCTTTCAA ATGCGGATTT	3360
	ACAAGTTTTT GATATAATTG ATCACTCGTG TTCAACTCAG TTGCATCACT TGTAACATTT	3420
30	AACACAACAG GTGCAGCAAT TCCGATTGCG ATAACCACAA CTAATAAAT ACTTAAATT	3480
	CTTTTTCGAT ATAGCGGGAT TTTCTTAAAA TTAACCTCCT CCCCATCTTT ATCTCCAAAA	3540
35	TACATATAAA AAAGTATCCT AAATAAACTG TACATTGCAA TTAGACTAGT AATAATCATT	3600
	AACGCTAGTC CAATATAATT GCCATTTTGC AATGCACCTT GGAAAATAAG TACTTTCCCC	3660
	GGAAAGCCAC TAAATGGAGG CACGCCGCCA ATAGCAAAAA TCATTATAAT AAACGCAACT	3720
40	CCAAATAAAG GTTCTTTTTT AGCTAAGCCA TTCAAATATT GATATTGTCG ATAGCCTGTA	3780
	ATGTAAACTA AACTACCAAT AATAAAAAAT AGCAATGTTT TTACAACAAT GTCATTTACC	3840
	AAATAAAATA TTGCACCATT AATACCTGCA AACGTGTTTG TTCCTAAACC TAAAATGATA	3900
45	AATCCTATTG AGATTATGAC TTGGTAAGCT GCAATCTTTT TAATATCTTT ATAAGCAATG	3960
	ACACCTATAG CGCCGATGAC CATAGTTATA GCAGCCATAG TTGCTAGCAA TGGATGTATG	4020
	AGATCATTAT GTTGATCAAA TAGTAAAGTG AAGAATCGAA TTAATGCATA GGCCCCTACT	4080
50	TTGGTCATTA ACGCTGCAAA TAATGCTGCA AGCTCAGTAT TTAACACAGC GTAGGCTTTG	4140
	GGTAGCCACA TAAAAAGGAC CAGCGCTGCT TTCGCACTAA ATGCGACTAA GAAGATTAAT	4200

55

	AAGTTTAATG TACCTACTGT TTTATAAAGT AAACCTATAC CTAATAAGAA TAGCCATGAA	4320
	CCAATAATAT TCAAGACAAC ATAAATAATT GCAGCACGTA ATTGTTCTAC AGATTGTCCA	4380
5	AGTGTAATGA GTACAAATGA CGCTAGTAAC ATAATTTCAA ACATGACGTA TAAATTAAAT	4440
	AAATCTGATG TTAGAAAAGA GCCTATCACG CCAACACTTA AAAATAATAT GAACGATGGC	4500
	AAGTGATAAC GATTTGCTTT ATGTTGCGCA CGCCCAAATC CGTATGCCAT AATTAAAGTA	4560
10	ATCACAAACG AAGCGGTTGT AACCATAATT AAACCTAAAG AATCTCCTAA AAACTGTATA	4620
	CCAAAGGGCG CTGACCATCC TCCAAAGTCT AGCGTAATTG GACGGTGACG CTGAACATAA	4680
	ATTAATAGCA TTAATGAAAT AATTGTGGTG ATAGTCATTG TACCTAAGTA TAAATATTTA	4740
15	GAAATACGAT CATTATTTTT TAAAAATACA AGGATTAAGG CACAAAGGAA TGGTAATAAC	4800
	ATTGGTAAAA TCAATAAGTT ACTTAGCATC ATCTTCCCCC CTTAGGCCTT CAATTTTCATC	4860
20	TTCTTTTGTT ACTTTATAAG TTCTATAAAC AAGTACAAGT AAAACGCAG TCATCCCAA	4920
	CCCTATAACT ATTGCAGTTA GTACAATAGC TTGTAACAAG GGATCAACAA ACAATTGGTT	4980
	TCCACCAGTT ATTAGTGGTT CTGATCTACT AGAACCATAC GTTCCCATAC TCATAATAAT	5040
25	GAGATTACCA GCATGAGTAT ATATTGAAAT TCCGATTACA ATACGAATTA AATTGATTGA	5100
	TAAAATCATA TATGTTCCCTA TAAACACTAA AAATCCTATA ACTAGTAATA ATATTAAATT	5160
	CATGATCGAC CTCGCTAAG CGACAACATC ACTGTGACAA TAACACCAAC AACTGAGAAT	5220
30	AAAATACCTA ATTCAAAAAG TGTTATTGTA CTTACATGAA TTTGTCCTAA AATTGGAAGT	5280
	ATCCAAGTTG TTTTCATATTG AGACAAAAT GGTTTTCCAA AAAACATAGG TATTATCGCA	5340
	GTAATAGATG ATACCAATGC TCCAATAATC ATTAATAATC TAAAATCAAT CGGTAACTT	5400
35	TCTAAAACCT CTTCAACATT AAAAGCCAGA AACATTAAAA TAAACGCTGA ACTAAATATT	5460
	AAACACCAA TAAACCCACC ACCAGGATTA TTATGACCTG CGAAGAAGAC ATAGAATCCG	5520
	AAAGTCAATA AAATAAATAC AACAAGTTTC GTGACCGTTC TTAACACGAC ATCATTCTCT	5580
40	TTTCATCTGT CCCCTCCGAT CTTGATAATT TAATAATGtg TAAATACCTA GCCCAGTAAT	5640
	AATTAACACT AATCCTTCAA ATAATGTATC TAATGCTCTA AAGTCACCAA GTATCGCATT	5700
45	TACAATATTT TTACCACCTG TTAGTTTGTC AGCTTTTAAA TAAAAGTCTG ATATTGATGA	5760
	TAAACCATCT GTTTGTTGTG TAATAAAAAT TAATGATACA ACAATAAGTG CCATCAAGAG	5820
	TGATACAGAA ATTTTAATTA TTTCTCTTTT TTTGTTAGCG TTAGATCTTG GCACGTTTGG	5880
50	TAATCTTGAA AAACCTGACAA TAAATAGTAT CGTCGTTATT GTTTCAACTA CTAGCTGAGT	5940
	CAATGCTAGA TCAGGGGCTT TCATTGCTAT AAAGAATAAG GTCACAACAA ATCCGATGAC	6000

55

	GACAGTTACG ATTGCTAATA TAATTTCTAA TGCCCCAAAT TCAGAAACAT GTAACGTATG	6120
	TACTTTAGGA AGTCCaATTC GAATATAACC ATATCCAATG ATAATCATAA ATATGCCTAA	6180
5	GGTCATAATA ATGTACTGGT TTAAACGATC TTGCATAACA CGTTTAAATC GCTTCGTAGC	6240
	AAACTTTTCA AAATGTGCGAT ATACCATCTC ATAGCTTTTT GAAACTGAAA TCTGTCTAAT	6300
	TTTACCTGTG AACACTTTTT TCCAATCTAC TTTGATTGCT AGTACACTAC CCAATAAAAT	6360
10	AATGATGATG GTTAAAAGAA GCGGTATGTT AAATCCATGC CATTGCGAAA CATGTGGTGC	6420
	CAATTGATCA ATTTGATGAT TACCACCTGA TACAGCTCTT AATGChAGAA CGATAATCCC	6480
15	CTTCCCAAAT ATATnTGGTA CAAAAAGAT TACAGGTACT AGCACCATTA aTATAAGAGA	6540
	TGGTAAACTA aACAACCATG GTTCGTGGAT ATTTTTTTTA GTAAAAACCT TAGAATCATA	6600
	TTTTGtCCAA AATACTTCTT TTACCATGTA TAGTGCATAT GTGAATGTAA AAACACTCGC	6660
20	AATAACACCA ACAACACGA TAGCTATCAT TGAAATCAAA CTAAATTGGG ATAATTGTCC	6720
	AGTTTGTGTT AATGCATCTA AAAACATTTT TTTACTTAAa AATCCATTTA AAAATGGTAC	6780
	TCCAGCCATA GATAGAGCCG CTATCGTCAT GACTAGATTG ATTTTAGGAA ATAGTTGACG	6840
25	CATTCCACTT AAAATTGTA TATCCCTTGA ACCTGCTTCA TGATCTAAAA TACCTACTCC	6900
	CATGAAAAGC GCACATTTAA AGATGGCATG ATTCATTAGA TGAAATAGcG CACCArATAA	6960
	TACmAATACA TAAATaGATG CTATTGCGTC TTGTTGGTGT TGAGCATATC CGCCACCTAT	7020
30	ACCCACCATA GCCATAATCA TCCCAAGTTG ACTGATTGTA GAGTACGCTA GGATACCTTT	7080
	TAAATCCCAT TGTTTTAAAG CTGTAATTGA ACCAAATAAC ATTGTTATTA AACCACAAAA	7140
	CGTAACGATA TATACGTACA TATTGCTAnG ACCTAATAAT GGTGTAAATC GAAGTAATAG	7200
35	AAnGATACCA GCTTTTACCA TCGTGGCTGA ATGTAAATAA GCACTTACAG GTGTAGGTGC	7260
	AGCCATTGCT CTAGGTAGCC AGTATGAAAT GGArATTGTG CTGATTTTGT AAATGCACCT	7320
	AATAAAAACA TAAAAATCAT AGGGATAAAC AATCCATGAT TCTTAATATG ATCTGCTTGT	7380
40	CCTAATATCT CTGTGATGTT ATTCGTTCCCT GTCATGATAT ACAGCATAAT AAAACCAACT	7440
	AATAACGCCA ATCCACCAAA TACTGTAATC ATAAATGATT GAATCGCACC AAATTGACTG	7500
45	TCACCATTGT TATACCAATA TGAnATCAAT AAAAATGATG ATmCACTCGT TAATTCCCAA	7560
	AAaATGTACA TCmATATCGT ATTGTCTGAT AATACaaTAC CAATCATACT GAACATAAAT	7620
	AACGTTAAAT AAAAATAAAA CCTTGGTAAA TTGTCTTTTC GAGAGGATAA ATATTGAGTT	7680
50	GCATAGAAGA ATACTGCAAT TCCAATAAGT GAAATAATAA GAGAAAACAT TAAACTTAAA	7740
	CCATCTAAAC GTAAATCTAA ATTAATATCT AATGTCTTAA TCCATGGAAT AGAGGTAGAA	7800

55

GGTGCAACCA ACGCTATGTA CCCGGCATAT TTAGCCAATG CTCTACGTTT AGACATTAGA 7920
 AGTATCATCG CCATAATCAC AAGTATAGCA ATTAATAAAT AAACCAAACT CATTATTAGC 7980
 5 CTCCTTTGTT TCTATAATTG TAATGAAATA TAAATACTAT GTTCACACTC ATTTTCTAAA 8040
 CCGATAAAAT TTAGTGTTTC AATAGCAGAT TGATGCCCTA AATACTTTTG AATGACTGGT 8100
 10 ATAAGTATAC CTTTTTGATA AGCATGATAT GCAAATGTCT TACGCAATGT CGTTAGTCCT 8160
 ACATTATCTA TACCAGCTTC AATTGATGCT TGGTGAATTA TTCGATATGC TTGCTGTCTA 8220
 GATAATACTT GATTGTTCG TAGTGATTGA AAAAGAACGT CTTCAATCGA AAGACTCCTG 8280
 15 TCCTCTATAT ATTGAAGTAG TTCTTTCGAT AATGTTTCTG GTAACCTAAT TTTAATCAA 8339

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

CCCGATTTTT TTACGTAATC TAATACATAC GGCAAATCA ACTTTAATCA AAAAAGACTC 60
 ATACACAATG CCTTTAAAGC ACATGTATGA GTCCTTTTGA GTAGTTTATA TCAAAAAATA 120
 30 GTTTAATGTA TAAATTAGTT TTTGTTTACA GATGCGTCGT AGATTGATTC TACAGCATCA 180
 CCTAAAGCTT TATCGAATTC TTCTTTAGAT TGATCAGCTC TTAAATCACT AGCTAATGCA 240
 35 CGTGAGAAAC TTGCGATAAG TTCAGCGTTA TCTTTAAGTA ATTCATTTCG TTTTCTCTG 300
 CTGTAACCAC CTGATAATAC AACGACACGA ACAACATTAG GATGTTTCAGC TAACTCTTTG 360
 TATAAGTTTG GTTCAGTAGG AATTGTTAAT TTCAACATTA CTAATTGATC AGCATTTAAG 420
 40 CTATCTAAAC CTTTTTAAAG TTCAGCTTTT AATACTTTTT CAATTCAGC TTTGTCTTTT 480
 GCATTAATAT TAACTTCTGG TTCGATAATT GGAACATAAC CTTTAGCAAT AATTGTGTTA 540
 GCAACTTCAA ATTGTTGTTT AACAACTCTT TTGATACCTT GTCATTT 588

(2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2841 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

	ATAGAGTnCT GGnACTTACT ATGACATATG GCGCTAGAAT GGCTGAGCCA GGTGAATTTA	60
	CAAAACGTGC CTTTTTAAAT GGTCGTATTG ATTTATCTCA AGCTGAAGCA GTTATGGACT	120
5	TTATTCGCTC GAAGACAGAT AGAGCTTCTA AAGTTGCGAT GAATCAAATT GAAGGTCGTC	180
	TAAGTGACTT AATCAAAAA CAACGTCAAT CTATATTAGA GATACTCGCT CAAGTGGAAG	240
	TGAATATTGA TTATCCTGAA TACGATGATG TTGAAGATGC GACTACTGAA TTTCTTTTAG	300
10	AGCAGTCTAA AGAAATCAAA CAGGAAATTA ATCGTTTATT AGATACCGGT GCGCAGGGTA	360
	AAATTATGCG TGAAGGTTTA TCTACAGTTA TTGTTGGTAA ACCAAACGTA GGTAAATCAT	420
	CGATGTTAAA TAATTTAATA CAAGATAATA AAGCGATTGT AACTGAGGTA GCAGGTACTA	480
15	CTAGAGATGT CTTAGAAGAG TACGTCAATG TTCGTGGCGT GCCATTAAAGA TTAGTTGATA	540
	CTGCTGGTAT ACGTGAGACA GAAGATATAG TAGAGAAGAT TGGTGTGAA CGCTCTAGAA	600
	AGGCTCTTAG CCAAGCAGAC TTAATTTTAT TTGTATTAAA CAATAACGAA GCATTGACwC	660
20	AAGAAGATTA CACATTATAT GAAGTGGTTA AAAATGAAGA TGTAATCGTA ATTGTTAATA	720
	AAATGGATTT AGAGCAAAAC ATAGATATTA ATGAAGTTAA AGATATGATA GGTGATACGC	780
25	CATTAATTCA AACTTCAATG TTAACAACAG AAGGTATTGA TGaATTAGAA ATACAAATTC	840
	gAGATTTGTT CTTTGGTGGA GAAGTACAAA ATCAAGATAT GACTTATGTT TCTAATTCAA	900
	GACATATTTT ATTATTAAAA CAAGCAAGAC AAACGATACA AGATGCGATT GATGCAGCAG	960
30	AATCTGGTGT GCCTATGGaT ATGGTACAAA TTGATTTAAC TAGAACTTGG GAAATATTAG	1020
	GAGAAATTAT TGGTGAGACT GCAAgTGATG AACTCATCGA TCAGTTATTC AGTCAATTCT	1080
	GCTTAGGTAA ATAGTAATTG AAATAGACGG AATACCGTCT TAAGAAGGCT AGTAAGATAT	1140
35	CAAATAAGGA GGTTTATATT GTGGTTCAAG AATATGATGT AATCGTTATA GGTGCGGGAC	1200
	ATGCAGGTGT AGAAGCAGGT TTAGCATCTG CAAGACGTGG TGCTAAAACA TTAATGCTAA	1260
	CAATAAATTT AGATAATATT GCATTTATGC CATGTAACCC ATCTGTAGGT GGACCAGCTA	1320
40	AAGGTATCGT TGTCGTGAA ATTGATGCTT TAGGTGGACA AATGGCAAAA ACAATCGATA	1380
	AAACACACAT TCAAATGAGA ATGTTAAATA CAGGTAAAGG ACCTGCTGTA AGAGCACTAA	1440
	GAGCGCAAgc AGaTAAAGTA CTTTATCAAC AAGAAATGAA ACGCGTGATT GAAGATGAAG	1500
45	AAAATTTGCA TATAATGCAA GGTATGGTAG ACGAACTTAT TATAGAAGAT AATGAAGTTA	1560
	AAGGTGTACG TACAAATATT GGTACAGAGT ATTTATCTAA AGCAGTAATT ATTACAACGG	1620
	GAACATTTTT ACGTGGTGAA ATCATTTTAG GTAATATGAA GTATTCAAGT GGACCAAATC	1680
50	ACCAATTACC ATCAATCACA TTATCAGACA ATTTAAGAGA ACTTGGTTTT GATATTGTTC	1740

55

AAATACAACC AGGTGACGAT GTAGGTCGTG CATTGAGCTT TGAAACAACA GAATATATAT 1860
 TAGATCAATT GCCATGTTGG CTAACGTATA CTAATGCTGA AACACACAAA GTTATCGATG 1920
 5 ATAATTTACA TCTATCTGCA ATGTATTCAG GGATGATTAA AGGAACCGGG CCACGTTATT 1980
 GCCCTTCAAT TGAAGATAAA TTTGTTTCGAT TTAATGATAA GCCGCGACAT CAACTTTTCT 2040
 TAGAGCCTGA AGGTGCTAAT ACAAATGAAG TATATGTGCA AGGATTGTCT ACAAGTCTTC 2100
 10 CTGAACATGT GCACGTCAAA TGTTAGAGAC GATACCAGGT CTTGAAAAAG CAGATATGAT 2160
 GCGTGCCGGC TACGCAATTG AATATGATGC GATTGTGCCA ACGCAGTTAT GGCCTACACT 2220
 TGAAACGAAA ATGATTAAAA ACTTATATAC TGCAGGTCAA ATTAATGGTA CATCTGGTTA 2280
 15 TGAAGAAGCA GCAGGACAAG GATTGATGGC AGGTATTAAAC GCTGCAGGTA AAGTGTTAAA 2340
 CACAGGCGAA AAGATATTAA GTCGTTTACA TGCATATATT GGTGTCTTAA TCGATGATCT 2400
 TGTAATAAAA GGTACTAATG AACCTTATCG TTTACTAACA TCACGTGCAG AATATCGTTT 2460
 20 GTTAcTACGT CATGATAATG CTGATTTGAG ATTGACGGAT ATGGGATATG AACTTGGTAT 2520
 GATTTCTGAA GAAAGATATG CACGTTTTAA TGAAAAACGT CAGCAAATTG ATGCGGAAAT 2580
 TAAGCGTTTA TCAGATATTC GTATTAAACC AAACGAACAT ACGCAAGCGA TTATTGAACA 2640
 25 ACATGGTGGT TCTCGCTTAA AAGATGGTAT TTTAGCTATC GATTTATTAC GCAGACCTGA 2700
 AATGACTTAC GATATAATTT TAGAACTTTT AGAAGAAGAA CATCAATTGA ATGCAGATGT 2760
 30 TGAAGAACAA GTAGAAATAC AAACAAAATA TGAAGGTTAT ATCAATAAAT CACTACAACA 2820
 AGTTGAGAAA GTTAAGCGTA T 2841

(2) INFORMATION FOR SEQ ID NO: 178:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3025 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

ATCTAATTTT AAACCCGGTG ATAAATTGCC AAGCGTGACG CAATTAAAAG AACGTTATCA 60
 45 AGTAAGTAAG AGTACTATCA TTAAAGCATT AGGCTTATTG GAACAAGATG GTTTGATCTA 120
 TCAAGCACAA GGCAGTGGTA TTTATGTGAG AAATATTGCT GATGCCAATC GTATCAACGT 180
 CTTTAAGACT AATGGTTTCT CTAAAAGTTT AGGTGAACAC CGAATGACAA GTAAGGTACT 240
 50 TGTTTTTAAG GAGATTGCAA CGCCACCTAA ATCTGTACAA GATGAGCTCC AATTAAATGC 300

55

	CGAATATTCT TATTATCATA AAGAAATCGT GAAATATTTA AATGATGATA TTGCTAAGGG	420
	CTCTATCTTC GACTATTTAG AATCAAACAT GAAACTTCGT ATTGGTTTTT CAGATATTTT	480
5	CTTTAATGTA GATCAACTCA CTTCAAGTGA AGCTTCATTA CTACAATTGT CTACAGGTGA	540
	ACCATGTTTA CGTTACCACC AGACTTTTTA TACAATGACT GGCAAACCCT TTGATTCATC	600
	TGACATCGTA TTTCATTATC GTCATGCACA GTTTTATATT CCTAGTAAAA AGTAATAAAT	660
10	ACATAAAAAC GTCTATATCC CAGTTATAAA CTGGAGTATA GACGTTTTTT TACGATAATA	720
	ACAATGGCTC AAATTGCTAT TATCTTGCTT AGGTTTTTCG TTTTAGAAGA ATATTGCTAC	780
	AAAGACAGGC ACAACTGCTA CAACAACTAC ACCAACTAAC ACTAAAGCTA TACTTGCCAT	840
15	TGATTCTTCT ACAGGTCCTA ATTCTTTGGC TGGTGCTACA CCTAATGTGT GACCACTTGT	900
	TCCAAGTGCT AATCCTCGGG CAATAGGGTT AGTAATTCGG AAAAGCTTTA AGAATTTATT	960
	ACCTAGGGCA TAAATAATGA CACCATTAA AATAACTGCT AATGATGTTA ATTCTTTTAT	1020
20	ACCACCGATA CCAGCTGATA CTGGTAACGC AATCGCTGTA GTTGCTGCTT GAGGTAACAT	1080
	TGATAAAATA ACATCATTGG CAAATTGTGC TAACTTCGCA AAAGTTAAAA TAATTAATAA	1140
25	CGCTACAAC GTACCGATAC CAATACCTCC GATGATACGA TGCCAATGTT TAACAAGCAC	1200
	TTCACGCTTT TTATATAACG GAATCGCAA ACAGATTGTT GCCGGTTCTA AGAAGAAGTA	1260
	AATAATGTCT CCACCTATTT TGTAAGTCTT ATACGGAATG CCTGTAAAT AGAGGAAGGC	1320
30	CACACCAAAT ACCATACTGA CAAATAGCGG TGCGAATAAG AAGAAACGAT TAGTTTTTTC	1380
	AAATAATATG GTCGCTAAGA AAAATGGTAT AACGGATAAC AGTATTCCGA AGTAAGGTGT	1440
	GTTTaGTGCT AAGTGGTTAA TCaTGAGCTT GTGCCTCCTC TATTTTGATC TTTTTGTGA	1500
35	CTTTGTCACC TTTAGATCTC GAAGTAACTT TCATAATAAT TTgTGTGACA TAGCCAGTAC	1560
	AAATaAGTAA TAGTATTGTT GAGACGATTA TTAGTCCAAT GATTAAAAAT GGTGCTTGGC	1620
	TAATGACACC TAAAGAGTTA ACAACTGAGA TACCGGCTGG TACGAAGAGT AAGCCAATGT	1680
40	TATTTGTTAG TGTCGTTCTT ACTTTTTCGA CTTCGCCTAA CTTAACAGCA CCAGTACATA	1740
	ATAATACAAA TAATAATACT AAACCGATTA CTGATGCAGG CATAGGAATT GGCATAAATG	1800
	ATTCAATTAT TTTCGATACA AAGAGTACTA AAGCAATTAC AATGACTTGG TGAAAAAGT	1860
45	GTGCTGGTTT TGATGCGTCT TTTTGTTGTT TCACGACCAT TGCCTCCTAC GTTTGATTTA	1920
	ACTAAAGTAT AGATGGCTCA CTTGATTTG CGTGATTTTT AGTCCGAAAT ACAAATATC	1980
50	ATAGGTAAAA TGCATAAAAA AAAGGATTAC TGTTAAAGTA ATCCTATCGA CGCTTTAAAA	2040
	TCTTTCATAA ATGAACGTCC AACTGTCATC TTGACACCAT TTGTCAATAT TACCATATAA	2100

55

	TGAATACGTA TAAAATAAGT GGGATTCAAT CGTTTTTCAT AACGATTCAA TGGCTCTGTT	2220
	GTTTCGTATT TATGATTCGT TGTATGTATG GTTGTAAATAC CATTATGTGT GCCAATCCCA	2280
5	ATAATATTTT GTTGCTTTAA CATGTGAATT TTATCGTCAA TTTCAACAGG TAAGCTTTGA	2340
	TCAAAATTCG CCGACATATC ATTCGCAATT GCACTTGCGT TATTATCATC TTTGGCTTTA	2400
	GTCGCACGCA CTTTATTGAC TGCTTGTTC AATACGTTTTT GACCAAACGG TTTCAAAATA	2460
10	TAGTCTGTCG CATTTAATTC AAATGCCTGT ACTGCGTATT GGTATGTGTC AGTTGCAAAA	2520
	ATAATCGCAG GTGGCTCTTT CATCTTTTGA ATCTTAGCTC CTAATTGAT CCCATTTTCA	2580
	TCCATTAAAT TGACATCTAA AAATATAATG TCATATTGAT TGATCAGTAG TGCTTCCAAT	2640
15	GTTTCTTTTA CATTCTTCTG CTCATTAATT TCTTCAAAC CACCAATTTT ATTTAATAAA	2700
	TATGTTAATT CATTACGTGC TAATGGCTCA TCATCTATGA TTAATGCTTT CATATTTATT	2760
	CCTCCTCTTG TCTTTCATAA GGAAGTACAC ACCAAAAAGT GGTACCGCTC GATGTCGATT	2820
20	CAAATTGTAA TGCTGCGGAT TTTCCAAATA ATCCTTTTAG GCGTAAGTTT AAATTTTCTA	2880
	AAGCACTACC AGTTCCAGAC TCTGATTCTA CAGATGTnTC TCCCaACAAA TGCATTTTAT	2940
25	CTTTAGAAAT ACCCTGACCA TTATCTTGTA CAATAATACG TACATGTGTT GCAGTTTCTT	3000
	TAATCACTGA CACGTCAATA TCGTT	3025

(2) INFORMATION FOR SEQ ID NO: 179:

30	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1689 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

	ACAGAATTTC ACAGCATTTT TAGATGAAAA AATAAGCCAG TCATAGCGTT GATTTAACAA	60
40	ATGAATATCA AAATTTAGTG GCTTTATATC AATAAAGGGT TTGTGAATAA TTGATACTAA	120
	ATCACTTTGC ATGTCATTTG TTTGTGTCAT AACTACAACCT GGCTTCATAT TTAAACGTCA	180
	CTCCATTATT TAATGTTGTT CATTAAAGCG TTTTATAATT TCATAAGCAC CTTGCTCTTT	240
45	TAATTTGTTA CTCACTGTTT TGCCTAACTC AACCGGATCT GTTCCGTTCA TTGTATATTC	300
	AAATCGTTCT TTACCATCTG GGGTCATAAT TAAACCTGTA AATTCGATTT CGTTTTGATC	360
50	TGAGATTGTA GCATATCCTG CAATTGGCAC CTGACAACTA CCATCCATTT CTGCTAAAAA	420
	CGTTCGTTCA GCAGTCACAC ATTTTGCAAC CTCATCATTA TGTACTTTGC TTAATAATGT	480

55

TAACAATGTA TCTCTATCAA GATAAGATGT TnCAATATCA TCTGACCAGC CCATTCTTCT 600
 TAAACCAGCT GCAGCTAAAA TAATCGCATC ATAATCTTCA GTTTGTAAGT TTTCTAATCG 660
 5 TGTATCTATA TTACCTCTAA TCCATTTAAT CTCTAAATTA GGATACTTAG ATAATATTTG 720
 TGCACCACGA CGTAATGAAC TAGTACCAAT AATACTGCCT TCTGGCAATT GGGATAGTGG 780
 10 TGTATGTGTT TTAGAAATAT ACGCATCAAA AGGTAATTCT CTATCAGGGA TACAACCTAA 840
 TGTAAACCT TCCGGAATTA CACTTGGTAC GTCTTTAAGC GAGTGTATTG CCATATCGAT 900
 ATTTTTTTCA AAAAGTTCAT GTTGTATTTC TTTAACAAAT AAGCCTTTGC CTCCGACTTT 960
 15 AGACAATTGT TTATCTACTA TACGATCGCC TTTCGTGACA ATTTCTTTAA TTTCAATTTT 1020
 TAGATTGGC TCGACAGCTT TTAATTTATC AATAAATTGC TGGCTTTGTG TTAAAGCTAA 1080
 TTTACyTCTT CTGGAGCCAA CGACTrATT ACGCATGTTT AATTCCTCCT AGGAACGGAT 1140
 20 TGCTCTAGAT TATTTTCTCA ATTCACAAAA TGTGTTGCAA AAAATAAATT AATCATATTT 1200
 AAGCAAAATA AAATAATGTT ATAGTATATT AAATATCTTG AATTCAACCA TTGTTGATT 1260
 CTAAGTAAAA TATAACTTCC ATATAATACT GTAATAATTG AAGAGAGTAT TACCTTCGGG 1320
 25 TCAATGAATA TACGTTCAAC AACTGAAATT ACACCCCACT GTGTACCTAA AATAATACTA 1380
 AATATGAGAA TTATCCACCC ACTTAACGTT GAGTAAAACA CAATTGATTC AAGTGTAGCA 1440
 ACGCTACCAA TTCTAAAGTA TTTTGTATCA AAACGTTTTT CCTTCAAATT ACGGTATTGC 1500
 30 ATGATATACA GTAATGCATT GACAAAAGCT AAGGCAAAGA AGACATAACT TAACACAGCT 1560
 AGACCGATAT GGACTAACAG TAACTCGTCT ACAACAGCAA TTTTCTGAAC CTTATTAGTA 1620
 TAATGTGTCG GTTGAAATGT ATTCATCCCT AAnAGTGTTA ACCCTATTAA ATTCCAAGGA 1680
 35 AAAACACAG 1689

(2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:
 40 (A) LENGTH: 1209 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

nTGGnTGGCT TTTCTATTG GACCAATGG ACCnTTTACC TGGCCnTTCC CAGGACACCC 60
 50 CGCTTGTGCC CACATTCCAA TCGGAAAAGG TGTATGTGGT ACAGCCGTTT CAGAACGTCG 120
 TACACAAATT GTAGCTGATG TTCATCAATT CGAAGGACAT ATCGCTTGTG ATGCTAATAG 180

55

5 CGATGCCCTT ATAACGGATC GATTTGATGA CAATGACAAa GAaCATCTTG AaGCAATTGT 300
 TAAAATTATT GAAAaGCAAC TCGCATAAAA GGACATCAGC ATTTTCAATA AAGTGTGAC 360
 AGTTAGCAGG AAAATGTTAC AATAATCTTT GTGTGAATTA ACGAAAGTAG CAGTTGTATA 420
 TTATTGAGCG CTATGTTGTT CCCAATGCCG ACGTGTACG TAACTGTCGC TATAAGGTGA 480
 10 AGACACATAA AACAAATATAT CTTAGTAAGC ATGCAACACT CTTTTTGTGTT TATTCATAAC 540
 AACAAAAAAG AATTAAAGGA GGAGTCTTAT TATGGCTCGA TTCAGAGGTT CAAACTGGAA 600
 AAAATCTCGT CGTTTAGGTA TCTCTTTAAG CGGTACTGGT AAAGAATTAG AAAAACGTCC 660
 15 TTACGCACCA GGACAACATG GTCCAAACCA ACGTAAAAAA TTATCAGAAT ATGGTTTACA 720
 ATTACGTGAA AAACAAAAAT TACGTTACTT ATATGGAATG ACTGAAAGAC AATTCCGTAA 780
 CACATTTGAC ATCGCTGGTA AAAAATTCGG TGTACACGGT GAAAACTTCA TGATCTTATT 840
 20 AGCAAGTCGT TTAGACGCTG TTGTTTATTC ATTAGGTTTA GCTCGTACTC GTCGTCAAGC 900
 ACGTCAATTA GTTAACCACG GTCATATCTT AGTAGATGGT AAACGTGTTG ATATTCCATC 960
 TTATTCTGTT AAACCTGGTC AAACAATTTC AGTTCGTGAA AAATCTCAAA AATTAAACAT 1020
 25 CATCGTTGAA TCAGTTGAAA TCAACAATTT CGTACCTGAG TACTTAAACT TTGATGCTGA 1080
 CAGCTTAACT GGTACTTTTC TACGTTTACC AGAACGTAGC GAATTACCTG CTGAAATTAA 1140
 CGAACAATTA ATCCGTTGAG TACTACTCAA GATAATACGG TCAATACCAA CACCCACAAT 1200
 30 TGTGGGTGT 1209

(2) INFORMATION FOR SEQ ID NO: 181:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 698 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

45 AAATCCCTTt GTtaAAGTsC AAAtTTTTTc AACrgCTTTA AtArGACCCA TATTACctTC 60
 TTGGATTAAA tCmAGGaATG AcATACCACG ACCaCGTATC TTTTAGCAAT ACTTACAAC 120
 AAACGTAAGT TCGCTTCTGC AAGTCTTGAT TTTGCTACTT CATCACCTTG TTCAATACGT 180
 TTGGCTAATT CGATTTCTTC TTGTGCACTT AATAAGTTAA CACGCCCAAT TTCTTTAAGG 240
 50 TACATACGAA CTGGGTCATT TATTTTAAAC CCTGGAGGGG CACTAAGATC ACTTGGATT 300
 AGTTTCTCGT CAGTATCTGA ACTATCTTTT TCATTAACTA GTGAAATATC ATTATCATTT 360

55

GCAATTTCTT CATGACTTAA ATGACCCTCT TTTTACCTT TTTCAATTAA TTGCTTCTTA 480
 ACATCTTCTA ATGTTAATGT CGGATCAATT GTTTGTITTT TAATTTTAAAC TGTGTTATCA 540
 5 GACATGAAAC GGCCTCCCGA TTTTAAATAT GAACATTGGA AATTTATTCA ATATTGCTAT 600
 TTTAAACGAA ATTCTTAATT AATTCCATCC ATATTTTAA TTTTATTTTA CAAATTGGGA 660
 10 ACTAAATCCC CAATATTAT TTTTCAATAG TGGTGGTT 698

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5147 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

ACTTGATGAT GTATACAATG TATTCAAGA ATATTATCAA AAAACATCTA ACATTAAGTT 60
 TTGTAGAATT CACAATTCTA GCTATTATCA CTCTCAAAA TAAAAACATC GTTCTTCTTA 120
 25 AAGATTTAAT TGAAACAATC CACCATAAAT ACCCTCAAAC TGTTAGAGCT CTCAATAATT 180
 TAAAAAGCA AGGCTATCTA ATAAAAAGAAC GCTCAACTGA AGATGAAAGA AAAATTTTAA 240
 TTCATATGGA TGACGCGCAG CAAGACCATG CTGAACAATT ATTAGCTCAA GTGAATCAAT 300
 30 TATTAGCAGA TAAAGATCAT TTACATCTTG TTTTGAATA ATATCTCTAT TACGCAAGTG 360
 TGCTGTATTC TAAAGTGCAC TTGTGTTTTC TATTTTTTAA TAAACCTCA GCACATAATG 420
 AACAACTTTC TATTTTCTAT ATCACTTAAA ACCATTTCCG AAATTAAACC TCAGCACATT 480
 35 CAAAGCCCCA CTTTATCTT AAAAATATTT TTAACTCAT ATGTATTAAA CCGCTTCAT 540
 TATAAAAAAT ATCTCTATAT TtTATCTGtT TtTATTAATC GAAATAGCGT GATTTTGCGG 600
 TTTTAAGCCT TTTACTTCCT GAATAAATCT TTCAGCAAAA TATTTATTTT ATAAGTTGTA 660
 40 AAACTTACCT TTAAATTTAA TTATAAATAT AGATTTTAGT ATTGCAATAC ATAATTCGTT 720
 ATATTATGAT GACTTTACAA ATACATACAG GGGGTATTAA TktGAAAAAG AAAACATtT 780
 ATTCAATTCG TAAACTAGGT GTAGGTATtG CATCTGTAAC TTTAGGTACA TTACTTATAT 840
 45 CTGGTGGCGT AACACCTGCT GCAAAtgctG CGCAACACGA TGAAGCTCAA CAAAATGCTT 900
 TTTATCAAGT CTTAAATATG CCTAACTTAA ATGCTGATCA ACGCAATGGT TTTATCCAAA 960
 50 GCCTTAAAGA TGATCCAAGC CAAAGTGCTA ACGTTTTAGG TGAAGCTCAA AACTTAATG 1020
 ACTCTCAAGC TCCAAAAGCT GATGCGCAAC AAAATAACTT CAACAAAGAT CAACAAAGCG 1080

	AAAGTCTTAA AGACGACCCA AGCCAAAGCA CTAACGTTTT AGGTGAAGCT AAAAAATTAA	1200
	ACGAATCTCA AGCACCGAAA GCTGATAACA ATTTCAACAA AGAACAACAA AATGCTTTCT	1260
5	ATGAAATCTT GAATATGCCT AACTTAAACG AAGAACAACG CAATGGTTTC ATCCAAAGCT	1320
	TAAAAGATGA CCCAAGCCAA AGTGCTAACC TATTGTCAGA AGCTAAAAAG TTAAATGAAT	1380
	CTCAAGCACC GAAAGCGGAT AACAAATTCA ACAAAGAACA ACAAATGCT TTCTATGAAA	1440
10	TCTTACATTT ACCTAACTTA AACGAAGAAC AACGCAATGG TTTCATCCAA AGCCTAAAAG	1500
	ATGACCCAAG CCAAAGCGCT AACCTTTTAG CAGAAGCTAA AAAGCTAAAT GATGCTCAAG	1560
	CACCAAAAGC TGACAACAAA TTCAACAAAG AACAAACAAA TGCTTTCTAT GAAATTTTAC	1620
15	ATTTACCTAA CTTAACTGAA GAACAACGTA ACGGCTTCAT CCAAAGCCTT AAAGACGATC	1680
	CTTCAGTGAG CAAAGAAATT TTAGCAGAAG CTAAAAAGCT AAACGATGCT CAAGCACCAA	1740
20	AAGAGGAAGA CAATAACAAG CCTGGCAAAG AAGACAATAA CAAGCCTGGC AAAGAAGACA	1800
	ACAACAAGCC TGGTAAAGAA GACAACAACA AGCCTGGTAA AGAAGACAAC AACAAAGCCTG	1860
	GCAAAGAAGA CGGCAACAAG CCTGGTAAAG AAGACAACAA AAAACCTGGT AAAGAAGATG	1920
25	GCAACAAGCC TGGTAAAGAA GACAACAAA AACCTGGTAA AGAAGACGGC AACAAAGCCTG	1980
	GCAAAGAAGA TGGCAACAAA CCTGGTAAAG AAGATGGTAA CGGAGTACAT GTCGTAAAC	2040
	CTGGTGATAC AGTAAATGAC ATTGCAAAAG CAAACGGCAC TACTGCTGAC AAAATTGCTG	2100
30	CAGATAACAA ATTAGCTGAT AAAACATGA TCAAACCTGG TCAAGAACTT GTTGTGATA	2160
	AGAAGCAACC AGCAAACCAT GCAGATGCTA ACAAAGCTCA AGCATTACCA GAAACTGGTG	2220
	AAGAAAATCC ATTCATCGGT ACAACTGTAT TTGGTGGATT ATCATTAGCC TTAGGTGCAG	2280
35	CGTTATTAGC TGGACGTCGT CGCGAACTAT AAAAACAAC AATACACAAC GATAGATATC	2340
	ATTTTATCCA AACCAATTTT AACTTATATA CGTTGATTAA CACATTCTTA TTTGAAATGA	2400
	TAAGAATCAT CTAAATGCAC GAGCAACATC TTTTGTGCT CAGTGCAATT TTTATTTTAC	2460
40	TTACTTTTCT AAACAACCTC TGAAACGCCT CAACACTTTC TACTCTGATT ACATATATGA	2520
	CATTTTTAGG CATTAAAAAA TCGAACTAGA CAAGATGCTC ATTGCATTTT GTACTAGTTC	2580
45	GATTCATGAA TAATTAGATT TAAAATGTCA TTTGAATCCA AGTGACAACA TTATTTATAT	2640
	TTAGAATATT AACGTIAGTA TAAACGTCCA AACACAAATA AAAGCAACAA ATATAATACT	2700
	GTATTTTAAC GTCATTTTTA ATAATGCAGA TTCTTCACCA ACTTTTTTAA CAGCTGCAGT	2760
50	CGCAATGGCA ATTGATTGTG GTGAAATAAG TTTGCTGCT ACACCACCTG CAGTGTTAGC	2820
	TGCCACAAGT AATGAACCGC TTGTTGAAAT TTGTTGTGCC ACTGTGCTT GAATAGGTGC	2880

55

	TGGAGAGAAT AATGGGAAAA TTGCTCCCGC TTTAGCAATA CCTTGTCCTAA TTGCTACAGT	3000
	CAAACCACCG TATGTCATAA CTTTAGCAAT AGCTAGGATA GCTGAAATTG TAAGGATCGG	3060
5	TAACCATAAT TCTTTAATTG CTTCGACCAA TAAAGCACCT GCACTTTTCC ATTTTAACTT	3120
	CGTAATTAAA ATTGTAATAA TTACTGTTAA TAAAATCGCT GTCCAGTTG CACCAATTAA	3180
	ATCGAGACGC AACGCAATTC CTTTAGGCGA TAAATCACTC ACAGTATTTG GAATTGGCAA	3240
10	TTTTATTACT AAACCTTCAA GTGCACCTCC AGGTTGGAAT AATTTTTTGA AGAATGGTGC	3300
	ACTCCATACT AATACAAAGG CAGTTAAAAT TACGAACGGA CTCCAAGCAA AGACAATTTT	3360
15	TTTAGGCGTT CGTTTTTGAA TTTTATGTTT AGACGCTTCC AATCTGAAAA TGTTTTTCGG	3420
	TTTAAATTTA CGACAAACAA ATGCTAACAC CACCATTGTT GCTAGTGATG GAATAATGTC	3480
	TGCTAGTTCT GGACCATGGA ATATTGTTAA TAATAATTGT AATCCAGTAT ATGTACCACT	3540
20	CACTGTAAAA ATGACAGGTA AAATTTCTTT AATACCTTTC ATACCATCTA CAATGAATAC	3600
	TAAAACAAAT GGAATAATAA AGTTTAAAAT TGGAAGTGTT AATGCTGAGT ATCTCGCAAC	3660
	ATCTAATGTT GTAACGCCTC CACTTAAGTT AAACGTATCA ATAATACTAA CTGGTAAACC	3720
25	AATTGCACCA AAGGCACCCG CCGCACCATT AGCAATTAAA CATAACATCG CTGCTTTTAA	3780
	TGGTTCAAAT CCAAGTTGAA TTAATAATAC TGCACAAATC GCAATTGGCA CACCAAATCC	3840
30	TGCTGCACCT TCTAAAAATG CGTTGAAACA AAATCCAATT AATAATAGTT GGATTCTTTG	3900
	GTCCACTGAA ATACTTGCAA TACTATCTTG AATAATAGAA AATTGTCCTG TTTTAATAGA	3960
	AACTTTATAT AACCAAACCTG CCATTAAAAAC GATATATCCT ATTGGGAAAA TACCGGCAAC	4020
35	AACGCCTTCT GTAATCGCAC CTGCTGATAC ACGCGCTGGT AATTCAAATA CAAATAAAGC	4080
	CACAAATCAAT GTAACAACCA AAGTTGTCAA TGCTGCATAA ATGCCTTTCA TTTTAAAAAC	4140
	GGTTAAGCAT AATAAAAAATA AAATAATAGG TACTGCTGCA ACTAAGGCTG ATAATCCGAC	4200
40	ATTATCGAAT GGATTACAG TAAGTAGTGT CATAATGACT CCCTCTCTTT ATATAAAATA	4260
	TTTATCATTG TGATTAATCT ACAACCTATT TCAACTTATA TTTTGCGATG ATCACATATT	4320
	TAAAATGTAA CACTCCTATA TGTGACAGGC AATCGAATTT TTACAAAAAG TTCACAAAAT	4380
45	ATACACAATA TTAACTATA ATAmATAATA TATCaTntTA ATTATAAATA CTAGATATTA	4440
	TTTATAATAA TCTCAGGAAT TCGCTTCAAA ACTGCATCAT GAGAGTTTAT ATTTTATTG	4500
50	AGAATCTCTC ATTTTATGAA TTGTAGGAAG TAAACAAAAT ATGACAAGCG TCAAACCAAT	4560
	GATAATGATA AATATCATAT TAAACCATAG TAAATTGAAT TGATGATGGT GTTGTATTTG	4620
	CCAAATTTCT AATACTGTGA AGATAGACAT ATAGCTCATA ATCTCTAAAT TTAACGTACT	4680

55

	AAATCGTTCA TAGTATCTAC CTGCAATGAA AAATATAAGC CAAATCACTA TAAATGCGCT	4800
	ATTAATCAAA AGCAGCACCC ATTTATCAGC AAAATTATCA GCATCCCCTG CTAAATTATA	4860
5	ATGAATAGGC ACTTTGGTTG GTAATTTTGG ATAGGTCACT ACTGTATAGC ACATCATAGC	4920
	TAAGTAAATA AGTAGACTTA ATATTGTAAA AGACCTGATT TTAGACATTC TATCGCCTcT	4980
	TcTTTACATT TTATGTATAA CACTCTGCCT ATTTTACCTT TTAATaCATT ACCCCAACGA	5040
10	TtAAaCAATA tGTAAaTGATA CTATAATTGC GTCAGGAGTA TCCGCTTGTT AAATGTGCAT	5100
	AGCTTATATT TAGCTGTTTA ACATGCCACA TAATGATTCG AATTATT	5147

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1312 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

25	CACTTACTTC CACCATTATC ATAACCTTAA AATGGATATA nTTCATCAAA CATTATCTAA	60
	AGGCGTCGCA CCTACACCAA CACCATCCAA CAATTAACCT ACAACTCTGC GATTACTTCT	120
	TCAGCAGCAA CTTTCACnTG CGTAATACAA TCAGGTAGTC CAACCGCTTC AAAAGATGCA	180
30	CCAGTTACTC TAAGTCGTGG ATATGTTTGT TTAATATGTG CTTGAATCTG TCTAATTTGT	240
	TGAATATGAC CGACATGGTA CTGTGGCATA CTTTTCGGCA AACGATTGAC AATTGTAAAT	300
35	TCAGGATCAC CTTTAAATGT CATCATTTGA CTTAAATCTC TACGTACAAT CGATACTAAT	360
	TCATTATCTG TATGATCATC AACCACAGTA TCACCTGGTT TACCTACATA CGCACGAATC	420
	AAAACCTTAC CTTCCGGTGT AGTAAATGGC CATTTTTTCG ATGTCCAAGT ACATGCCGTA	480
40	ATGTCTGTAT CACTCGTTCT CGCAATTACG AAGCCAGTAC CATCATGGGT ATTTTCAATG	540
	TCTTTTTTCAT CAAATGCCAA TACAACAGTT GCAACAGTCG TACTATCCAT CGTTTTAAAG	600
	TAATCAAATG CTGGATCTTG TCCGAACCAA TTTAAAAACA CTTGATGTGG TGTCGTTACT	660
45	AATACGCCAT CATACTTC TTCTAGTTGA TCATTGTAAA CAATTTTATA TTGTTTTTGA	720
	GATGTAATTA TATCATCCAC TGACGTATTG TAGCGTATTG TCACACCTTT ATTTTAAACA	780
50	TCTTGTTCTA ATGCTTCAAT AAATGAGCTT AAACCATGCT TAAATTGTTT GAATTGTCCT	840
	TTCGGTGCGC CAGGATATAA TTGTCTTTGT TTCAGACGCT TATTTTTCTC ATCCTTCATA	900
	CCTTTTATCA GACTTCCGAA TGCCTCTTCT TTTTCTTTAA AATTAGGAAA CGTACTCATC	960

TCAAGTACCT CATTACCTAA TCTTGCTCTG AAAAATGCAC CAACAGAAAT GTCACCATCC 1080
 TGCATTTGAG TAGGTTTTTT TAATAAATCA AACCTGCTC TTAATTTACC AAGTGGCGAT 1140
 5 ATTAATTTTG TAGTAACAAA TGGTTTAATA TCTGTTGGAA TACCCATAAT TGAACCACCT 1200
 GGAATCGGAT ATAATTTATT TTTTCGAAAA ATATATGATT GTCCAGTCGT ATTTGTAACA 1260
 ATATCTTGTT CTAATCCAAT ATCTTTCGCT AATTCTGTCA TAATCGTTTT TC 1312

10 (2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6157 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

TTTTACAATA AAAATATGAT ATACTACTTG TCGTATATAA GGAACGGAGG ACAATTTATG 60
 CATACATTTT TAATCGTATT ATTAATCATT GATTGTATTG CATTAAATAAC TGTGTACTA 120
 25 CTCCAAGAAG GTAAAGCAG TGGACTTTCA GGTGCCATCA GTGGTGGTGC TGAGCAGTTA 180
 TTCGGTAAAC AAAACAACG TGGCGTCGAT TTATTCTTAA ATAGATTAAAC AATTATTTTA 240
 TCAATATTAT TTTTGTACT TATGATTGTC ATAAGTTATC TTGGTATGTA AGGTCCGGCG 300
 30 ATGTAATGT CGGGCTTTTT TATTATAAT TAAGAATGTA ATAGTTTAAAC AATAAGCTAT 360
 GTAAATATA TAGCCTAGTT AAGTATGCAA AGGGAGCGTT AGATTTATGC AGATAAAATT 420
 35 ACCAAAACCT TTCTTTTTTG AGGAAGGTAA ACGTGCCGTG TTATTACTAC ATGGTTTTAC 480
 AGGCAATTCT TCTGATGTTT GTCAATTAGG TCGATTTTTA CAAAAGAAAG GTTATACATC 540
 ATATGCACCG CAATATGAAG GCCACGCGGC ACCACCAGAT GAAATACTGA AATCTAGTCC 600
 40 TTTGTTTGG TTAAAGATG CGTTAGATGG TTATGATTAT CTTGTTGAAC AAGGTTATGA 660
 TGAAATTGTT GTTGCTGGTC TATCATTAGG TGGGGATTTT GCTTTAAAAT TAAGCTTAAA 720
 TAGAGATGTA AAGGGTATTG TAACGATGTG TGCTCCTATG GGTGGCAAAA CTGAAGGTGC 780
 45 CATTTATGAA GGCTTTTTAG AATATGCACG CAATTTTAAA AAGTATGAAG GTAAAGATCA 840
 AGAGACTATT GATAATGAAA TGGATCATTT TAAACCAACT GAAACTTTAA AAGAACTAAG 900
 50 TGAAGCATTG GATACGATTA AAGAGCAAGT TGATGAAGTG TTGGATCCTA TTTTAGTGAT 960
 TCAAGCAGAA AACGACAATA TGATTGATCC ACAATCCGCA AATTATATAT ATGACCATGT 1020
 AGATTCTGAT GACAAAAATA TCAAGTGGTA CAGTGAATCT GGACATGTTA TTACGATTGA 1080

	AGAATAAAAA	GAGATTTTAA	CATTAGAAAAG	GAGGGGCATA	ATGAATTTAA	AGCAATCTAT	1200
	AGAAGAGATT	ATTAATCAAC	CTGAATATGA	ACCTATGTCA	GTGTCAGATT	TTCAAGATGC	1260
5	ATTAGGTTTA	AGCAGTGCCG	ACTCGTTTAG	AGATTTAATT	AAGGTGCTTG	TGGAGTTAGA	1320
	ACAATCAGGA	TTAATCGAAC	GTACAAAAAC	AGACAGATAC	CAAAAAAAGC	ATAGTTATAG	1380
	AGGTCAATCA	AAATTGATAA	AAGGAACGTT	AAGTCAAAAT	AAAAAAGGCT	TTGCATTCTT	1440
10	AAGACCTGAA	GATGAGGATA	TGGAAGATAT	ATTTATTCCC	CCGACGAAAA	TTAATCGTGC	1500
	CTTGGATGGA	GATACTGTTA	TTGTAGAAAT	CCATCAATCA	AAAGGTGAAC	ATAAAGGTAA	1560
15	AATCGAAGGG	GAAGTTAAGT	CGATTGAGAA	GCATTCTGTA	ACTCAAGTTG	TTGGTACGTA	1620
	TAGTGAAGCT	AGACATTTTG	GCTTTGTTAT	TCCGGATGAT	AAACGTATTA	TGCAAGATAT	1680
	TTTCATTCTT	AAAGGTCAAA	GTTTAGGCGC	AGTCGATGGT	CATAAGGTAC	TTGTACAAAT	1740
20	TACTAAGTAT	GCTGATGGTT	CAGATAATCC	AGAAGGACAT	ATTTCTGCTA	TTTTAGGACA	1800
	TAAAAATGAT	CCTGGCGTAG	ATATTTTATC	TATTATCTAT	CAACATGGCA	TAGAAATTGA	1860
	ATTTCTGAT	GAAGTGTTAC	AAGAAGCTGA	AGCAGTACCT	GATCATATTG	AAAATACTGA	1920
25	AATTAAAGGC	CGTCATGATT	TACGTGATGA	ATTGACAATC	ACAATTGATG	GTGCTGATGC	1980
	TAAAGACTTA	GATGACGCAA	TTAGTGTTAA	AAAGTTAGCG	AACGGTAATA	CGCAATTAAC	2040
30	TGTAAGTATT	GCTGATGTCA	GCTATTATGT	AACAGAAGGT	TCTGCATTGG	ATAAAGAGGC	2100
	ATATGATAGA	GCGACAAGTG	TATATCTTGT	TGACCGTGTA	ATTCCAATGA	TTCCACATCG	2160
	ATTAAGTAAT	GGTATTTGTT	CATTGAATCC	TAATGTTGAT	CGTTTAACTC	TAAGCTGTCG	2220
35	CATGGAAATC	GATGCTAGTG	GTGCGTTTGT	TAAACATGAA	ATTTTGTGATA	GTGTTATACA	2280
	TTCTGATTAT	CGAATGACGT	ATGATGCGGT	AAATCAGATT	ATTACTGAAA	AGGATCCTAA	2340
	CATTGCGGAA	CAATATAATG	AAATTACGCC	TATGCTAGAT	TTAGCACAAG	ATTTATCTAA	2400
40	TCGTTTGATT	CAAATGAGAA	AACGACGTGG	TGAAATCGAT	TTTGATATTA	GTGAAGCAAA	2460
	AGTATTAGTT	AACGAAGACG	GTATACCAAC	AGATGTTCAA	TTAAGACAAC	GTGGCGAGGG	2520
45	TGAACGTCTA	ATTGAATCAT	TTATGTTAAT	TGCAAATGAA	ACAGTTGCTG	AACATTTTAG	2580
	TAAGTTAGAT	GTACCTTTTA	TTTACCGAGT	GCATGAGCAA	CCTAAATCAG	ATCGCTTAAG	2640
	ACAATTCTTT	GATTTTATTA	CAAACCTTGG	CATCATGATT	AAGGGTACTG	GCGAAGATAT	2700
50	TCATCCAACA	ACACTTCAAA	AGGTTCAAGA	AGAAGTAGAA	GGTCGACCTG	AACAAATGGT	2760
	CATTTCAACA	ATGATGTTGC	GTTCAATGCA	ACAAGCGCAT	TATGATGATG	TGAACTTGGG	2820
	ACATTTTGGC	TTATCAGCTG	AATATTATAC	GCATTTTACA	TCACCAATTA	GACGTTATCC	2880

55

	AGAAGTGAAG CGTTGGGAAG ACAAATTGCC TGAGTTAGCT GAACATACTT CTAAACGTGA	3000
	ACGTCGTGCT ATTGAGGCAG AACGTGATAC TGATGaATTG AAAAAAGCAG AATATATGAT	3060
5	TCAACATATT GGTGATGAAT TTGAAGGTAT TGTCAGCTCA GTAGCTAACT TCGGTATGTT	3120
	CATTGAATTG CCAAATACGA TAGAAGGTAT GGTTCATATT GCGAATATGA CTGATGATTA	3180
	TTACCGTTTT GAAGAGCGTC AAATGGCATT AATTGGTGAG CGTCAAGCTA AAGTATTTAG	3240
10	AATTGGTGAC ACAGTTAAGG TTAAAGTGAC GCATGTTGAT GTAGATGAAC GATTAATTGA	3300
	TTTTCAAATT GTAGGTATGC CTTTACCGAA AAATGATCGA TCACAGCGCC CAGCGCGAGG	3360
	TAAGACAATT CAAGCCAAAA CGCGTGGTAA ATCATTAGAT AAATCAAAAT CTGATGATAA	3420
15	GGGTCGTAAG AAAAAAGGTA AGCAACGTAA AGGTAAAAAC CAACGTAATA ATGATAAATC	3480
	AGGTAATAGT AAGCATAAGC CATTTTATAA AGATAAAAGT GTGAAAAAGA AAGCACGTCG	3540
20	TAAGAAAAAA TAAGCAGCAA TGAGGTGAGT ATGAATGGCT AAGAAGAAAT CACCAGGTAC	3600
	ATTAGCGGAA AATCGTAAGG CAAGACATGA TTATAATATT GAAGATACGA TTGAAGCGGG	3660
	AATTGTATTG CAAGGCACAG AAATAAAATC AATTCGCCGA GGTAGTGCTA ACCTTAAAGA	3720
25	TAGTTATGCG CAAGTTAAAA ACGGTGAAAT GTATTTGAAT AATATGCATA TAGCACCATA	3780
	CGAAGAAGGG AATCGTTTTA ATCAGATCC TCTTCGTTCT CGAAAATTAT TATTGCACAA	3840
	GCGTGAAATC ATTAAATTGG GTGATCAAAC ACGTGAGATT GGTATTTCGA TTGTGCCGTT	3900
30	AAAGCTTTAT TTGAAGCATG GACATTGTAA AGTATTACTT GGTGTtGCAC GAGGTAAGAA	3960
	AAAATATGAT AAACGTCAAG CTTTGAAAGA AAAAGCAGTC AAACGAGATG TTGCGCGCGA	4020
35	TATGAAAGCC CGTTATTAAG CGATTTAGTT GCTTAATCGG GCTATATTTG ATATAGTTAT	4080
	ATGTGCTTTT GTAAATTACA AAAGTATGAT TTGTTTGATT TATTATTTG GGGACGTTCA	4140
	TGGATTCGAC AGGGGTCCCC CGAGCTCATT AAGCGTGTG GAGGGTTGTC TTCGTCATCA	4200
40	ACACACACAG TTTATAATAA CTGGCAAATC AAACAATAAT TTCGCAGTAG CTGCCTAATC	4260
	GCACTCTGCA TCGCCTAACA GCATTTCTTA TGTGCTGTTA ACGCGATTCA ACCTTAATAG	4320
	GATATGCTAA AACTGCCGT TTGAAGTCTG TTTAGAAGAA ACTTAATCAA ACTAGCATCA	4380
45	TGTGTTTGT TTATCACTTT TCATGATGCG AAACCTATCG ATAACTACA CACGTAGAAA	4440
	GATGTGTATC AGGACCTTTG GACGCGGGTT CAAATCCCGC CGTCTCCATA TTTGTAGCCT	4500
50	ACAGCCTTTG TGGTTGTGGG CTTTTTTATT TTGTGTTTTT CAGGGGATAA TGCATTGCAG	4560
	AATTTGTTGT GAGTATTGAT ATAGCAGTGT TTGTATAGGT GTTTATTTGA TGGAGGAAAG	4620
	AGTAATAAGT GATTATGAAT TAGTTTTTGA GATATAAGGG GACAGTGATG TGTGTCAAAT	4680

55

TTATACGCAA AAAATTCTCC ATGTTATATA TGTCAATATA AAAATGTGAA TCGTCTACAC 4800
 TTAATTGGAT AAATGGCTAC TGAAAAAGAA CTTTTCATTT TTGTTACGTC ACTAAGTGGG 4860
 5 TGTAGTTATA AAGAGATGAG CCGAGTTTTG ATATTTTCAT TAGAATCAAT ATGCCTATTA 4920
 ACACAATCAG CAATAGTTGA CGAGACGGAA ATAAAAGAAG TCGTAGTTAA GAAATGCATT 4980
 TCACAACATA CCATTGTAGC CATTTTTTATT GTTTTGGATG ATAAACTCTT TTTGGAATTT 5040
 10 TTAGTTTTTA TAATTGCAA CTACACTACT TCTTTTACTA ATATTAATGT CTAAGTAATC 5100
 GATAAAAAAT TTTCCATTGA ATAAATGAGA AGTTAAAAAC TTTACTTAAC CTTTCyCATT 5160
 GCATTTTCCT ATTCACGATT TTAAGAACCC AACATACTAC AAACGAATTT TAAAAGGCGA 5220
 15 GAGTAAAGCT TACTTGTTTA TTATACATAT TTAAATCCA AGAGTCAGAA CAGACTACTC 5280
 CTCTTTATAA CTATAAAAAA TAGCTATGAA AAAATCTATC GTCATAGATT CCTTCATAGC 5340
 20 TAATCTTAGT ATGTTTATTT TTATTTTAGG ATGCTATTTA TCAACTCAAC ATATAACTCA 5400
 CTATTTTTAT AACCTTCTAA TATATCATTa ACTTGTCTAA TAGGTATTTc TGGTACTTCT 5460
 CTAATGTTTT CCAATTTTGT TTTAAATTGT TTTTTTGTTA TTTGCTCTTT ATTGTAGCC 5520
 25 AATTGGAACA AGTAAGAATC TAGCATATTA ATTTCTTTAT ATGAATACAT ATATCTTAAT 5580
 AACACTAAAT CTCTAGTTTT TAAGTTAGGC GCTAGTTCTT CTTGTAATTG TTCTATTGAT 5640
 TGTyTCATTA ATAACAATCT CATTTCTAAT TCTTCATTAT TCATTTTATC ACACTCTTtT 5700
 30 TATATTAATG CTTGACCAAC TTGGGAAACC CAAAACCCTA TGCTTCTTGC AGTAGAATCT 5760
 TTAATACCAG TTCCCATCAA TGCTTG TGAA ACTTGACCTT GTACATTTCC CCATGTAGCC 5820
 35 TCTTCTTGTT TTAATGCATT ATTCAATGCG GGATTTACAA ATTTATCCCA TCTTTTTTTT 5880
 ATGATTTTCC GGCACGGGGA CTGATTTCTT TAACACCAAT AAACACAGAT TTTTATTTTT 5940
 TAATCATAGC TTTATAGTAT CATGTTGGCT AAGCTATAAA TAAGTCAGTT TCTCTAAAAA 6000
 40 TTAAATAACT GAATGTAAGA CAATCAACAA WCCAAATTTA TACTTCATCT AAACCACTGT 6060
 GGTGTCATC TTTTGTCTT TCTTTTTCTT TCTCTCGTTC TTGTTCTTTT TTGTACTCTT 6120
 CTTCAAATTC TTTTCTTTT TTTTCTACTT CTTCTCT 6157

(2) INFORMATION FOR SEQ ID NO: 185:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 884 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

CATTGTAT TCTGAGTAGC CAATTGGCA AAGATGAACA AACGTCTGAA CAAACGTATC 60
 AAGTTGCAGT CGCATTAGAG TTAATTCATA TGGCAACACT TGTTTCATGAT GACGTTATTG 120
 5 ATAAAAGCGA CAAGCGTCGA GGCAAGTTAA CCATATCAAA GAAATGGGAT CAGACAACCTG 180
 CTATTTTAAC TGGGAATTTT TTATTGGCAT TAGGACTTGA AACTTAATG GCCGTTAAAG 240
 ATAATCGTGT ACATCAATTG ATATCTGAAT CTATCGTTGA TGTTTGTAGA GGGGAACTTT 300
 10 TCCAATTTCA AGACCAATTT AACAGTCAAC AGACAATTAT TAATTATTTA CGACGTATCA 360
 ATCGCAAAAC AGCACTGTTA ATTCAAATAT CAACTGAAGT TGGTGAATT ACTTCTCAAT 420
 CTGATAAAGA GACTGTACGA AAATTGAAAA TGATTGGTCA TTATATAGGT ATGAGCTTCC 480
 15 AAATCATTGA TGATGTATTA GACTTCACAA GTACCGAAAA GAAATTAGGT AAGCCGGTCG 540
 GAAGTGATTT GCTTAATGGT CATATTACGT TACCGATTTT ATTAGAAATG CGTAAAAATC 600
 20 CAGACTTCAA ATTGAAAATC GAACAGTTAC GTCGTGATAG TGAACGCAAA GAATTTGAAG 660
 AATGTATCCA AATCATTAGA AAATCTGACA GCATCGATGA GGCTAAGGCA GTAAGTTCTGA 720
 AGTATTTAAG TAAAGCYTTG AATTTGATTT CyGaGTTACC aGATGGACaT CCGaGatCAC 780
 25 TACyTTTAAG TTTGACGAAA AAAATGGGTT CAaAaAACAC GTAGTATTTA TGnAAAAGTA 840
 TTGAAAGCGC TTTACCAACC TGTTAATATA TAATAGTAAT ATAC 884

(2) INFORMATION FOR SEQ ID NO: 186:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6876 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:
 40 AATTTTCATCT GCTCGTGCAA AATCTTTGTT TTTCTTGCT TCATTACGCT CTTCGATTAA 60
 TTTTTCACAA TCTTCATCCA ATAATTCATC TGCATTTTGA GATTTTAACG GTACACCTAA 120
 AACATCGCTG AAAATTTGAT AAAGTGCTTT AAATTTATCA ATTACTTCTG TTGATGTTGT 180
 45 GTTCTCTAGT ACATATTTAT TCGCAAGTKT TGCTAAATCA TACCAAGCTG TAATGTCATT 240
 AGCTGTATTA AAATCATCAT TCATAACTGT TTCAAAACGA TTTAAAATCG CATCAATTTG 300
 ATCAATATAT GTCTGTTGAT TTTCAATATT AGTAGCAATT TGTGCGCGCT CTTCAAATTA 360
 50 TTGATAACTA TTGCGAATAC GCTCTAGTcC aCTACGTGCT GATTCTACCA ATTCTAGATT 420
 ATAGTTAATT GGGCTTCTAT AATGTACGCT AATCATAAAG AATCTTAGTA CATCTGGATC 480

55

	ATTATCAATA TTAATGAAAC CATTATGCAT CCAATAATTA GCAAATGGCG CATGATTATG	600
	TGCTTCTGAT TGTGCTATTT CATTTCATG ATGTGGAAAT TGTAAATCTG AACCACCCGC	660
5	ATGTATATCA ATTGTAGGTC CTAGCTCATG AAATGCCATT ACAGAACATT CTATATGCCA	720
	TCCTGGTCTA CCTTCACCAA ATGGGCTATC CCAACTAATC TCGCCAGGTC CGCTTTTTTC	780
10	CACAATGTAA AATCAAGTGC ATCTTCTTTA TGCTCTCCTG CATCTATACG AGCACCCACT	840
	TTTAAGTCAT CTATGGATTG ATGACTTAAT TTACCATAAC CTTCAAATTT ACGTGTTCTA	900
	AAGTAAACAT CGCCACCACT TTCATATGCA TAACCTTGAT CCACCAAATC TTTAATAAAT	960
15	TGAATAATGT CATCCATATG GTCCATTACC CTTGGATTGG AAGTCGCTTT TCTAACATTT	1020
	AACGCACCAA CATCTTCATG AAAAGCAGCG ATATATTTTT CTGCAATTTT GGGAACAGAC	1080
	TGATTTAATT CTTGAGAACG TTTAATTAAT TTATCATCTA CGTCTGTAAA ATTTGATACA	1140
20	TATTCTACAT TATATCCTTG GTATTCAAAG TAACGTCTCA CTACGTCATA ATTAATTGCw	1200
	GGTCTTGCGT TACCAATATG AATGTAGTTA TATACAGTAG GACCACATAC ATACATTTTT	1260
	ACTTTCCCTG GTTCTATAGG CTTGAACACT TCTTTTTGAC GTGTAAGCGT ATTATATAAT	1320
25	GTAATCATCT TGAATCTCTC CATTCCTAGT CTTTTCAAGT TGTCGTTCTA AATGCTTAAT	1380
	TTGTTCATAA ATTGGATCAG GTAGATGGCG ATGATCAAAT GTTTTTCCAA CTCGAACACC	1440
30	ATCTTGCTTA ACAATATGTC CTGGTATACC AACAACCGTT GAATAACTTG GAACTGATTG	1500
	TAAAACAACT GAATTTGCAC CAATATTTAC ATTTGAATTT ATTTTAATAT TTCCTAAAAC	1560
	TTTCGCACCG GCTGCTATTA AAACATTGTC TCCTATATCT GGGTGTCTTT TCCCTCTTTC	1620
35	TTTCCCTGTC CCACCAAGTG TCACGCCTTG ATAGATTGTC ACATTATCAC CAATTGTACA	1680
	TGTTTTCTCT ATTACAACGC CCATACCATG ATCTATAAAT AGACGCTTTC CAATTTTAGC	1740
	ACCTGGATGG ATTTCTATAC CTGTGAAAAA TCTTGAAATT TGAGATATCG CGCGTGCTGC	1800
40	AACATATTTT TTTTGGTTGT ATAACCTATG TGCAATCAAA TGACTCCAAA CTGCATGTAA	1860
	ACCTGCATAC GTTGTAATGA CTTCTAATGT TGAACGTGCC GCTGGATCCT GCTCAAATAC	1920
	CATTTTTATA TCGTCTCTCA TTCTTTTTAA CAAGATCATT TCCTCCTCAA TGATTGAACT	1980
45	ACGTAAATAC ATAATTGAAG TACCTGCGAA ATTAAATATC AAAAAAGCAC CACTAACATA	2040
	CAAATTGTAT TGTTAGAGGC GCTTCCGCAC GGTTCCACTC TGAATTTAGC GAATAACATT	2100
50	AATAATATTG CGGGCGCTTC CAAATTATCA AGGAACTAA GTCAACTTAA TGCTCATCAC	2160
	TCTCATTATA TATTTAATTC ATTTTACGAA GGTGCATTCA TTAATTTCTA CGTTGTACTC	2220
	ACAGCAACCG TACACTCTCT GCATCGTATA AATTTAATTA CTAATCCTTC GTTTTATATA	2280

55

	ATAAAATTCA AGTATATACT ACCTTGATCT TGTCTATTTT ATTACTTATA TTGTTTTAAA	2400
	CGGTTTAGCA CTTTTTCTTT ACCAAGTACT TCAATTGTAT TTGGTAATTC AGGACCATGC	2460
5	ATTTGGCCTG TTACAGCAAC ACGAATAGGC ATAAATAATT GCTTGCCCTT TATTCCTGTT	2520
	TCTTTTTGAA CTTCTTTAAT TGTCTTTTTA ATTTCAGCCG CTTCAAATGG TTCAAGTGCT	2580
10	TCTAATTAC TGAATAAGTG CGTCATTAAC TCTGGTACTT GCTCTCCATT AATCACTGT	2640
	TGTTCTTCTT CACCAAGAGC TGGCATTCTT TTAAAGAACA TTTCTGATAA AGGTACAATT	2700
	TCACCGGCAT AACTCATTTT TTTTGTATAA AGCGCAATTA ATTTGCGTCC CCAAGATAAA	2760
15	TCCTCTTCTG ACGGCACCTC AGGAATCAAA TTTGCTTTAA TTAAATGAGG TAATGCTAAT	2820
	TGGAATACTG TTTCAGTATC TTTTGTTC ATATATTGGT TATTAACCCA TGCTAATTTT	2880
	TGCTTATCGA AAAATGCTGG TGATTTTGAC AAACGCTTTT CATCAAAGAT TTTGATAAAT	2940
20	TCTTCTTTAG AAAAGATTTT TTCTTCACCT TCAGGAGACC AACCTAATAA CGCAATAAAA	3000
	TTAAATAACG CTTCAGGTAA ATAACCTAAG TCACGATATT GCTCAATAAA TTGTAAAATT	3060
	TGCCCATCAC GTTTACTTAA CTTTTTACGT TCTTCATTAA CAATTAATGA CATATGACCA	3120
25	AAACGAGGTG GCTCCCAGCC AAATGCTTCA TAAATCATAA TTTGTTTAGG CGTGTGAA	3180
	ATATGATCAT CACCACGAAT TACATCTGAA ATTTGCATGT AATGATCATC TATAGCTACT	3240
30	GCAAAATTGT ACGTTGGAAT GCCATCTTTT TTTACGATAA CCCAGTCACC AATACCATT	3300
	GAATCAAATG AAATATTTCC TTTTACCATA TCATCAAATG AATACGTTTG GTTTTGAGGT	3360
	ACTCGGAAAC GAATTGATGG TTGGCGTCCT TCTGCTTCAA ATTGTTGACG TTGTTCTTCA	3420
35	GTCAAATGCG CATGTTGACC ACCATAGCGA GGCATTTTAC CACGAGCGAT TTGCGCTTCA	3480
	CGTTCAGCTT CTAATTCTTC TTCTGTCATA TAGCATTTAT ATGCTTTATC TTCTGCTAGT	3540
	AAC T GATCTA TTAATGGTTG GTAGATATGT TGACGTTTCA ATTGACGATA TGGTCCGTAG	3600
40	CCATTGTCTT TATCTACAGA CTCATCCCAA TCTAATCCTA ACCATTTAAG ATTATCAAAT	3660
	TGTGATGTTT CTCCATCTTC TAAATTACGT TTTTATCAG TATCTTCAAT TCGAATCACA	3720
	AAATCTCCGT TGTAATGTTT AGCATAACAAG TAATTGAATA ATGCTGTTCT TGCATTACCA	3780
45	ATATGAAGAT ACCCAGTTGG ACTTGGTGCA TATCTTACTC TTATACGATC GCTCATTTTT	3840
	TTCACTCCTA AATTAAATAT CAGATTTTCA AGTTAGTTCA TATAAATTGT TCATTTGCTA	3900
50	TCTTCGACCG TCATAACAAA TGTCTAACTC GTCTTATTGT TAAAACGAAA CAATGCTTTT	3960
	TAACATGACC TTAAATAAAT TTCATTGTTT AATCATAACA TAATCCCTG GGTAAATATGC	4020
	TTAAATTTTA AATAGAAAGC TGTTGTTTTT TCAACACTTT AAAAAAGCTA TCCCTAAGAA	4080

55

	TTAAACTTCA AATTAACAT TCAAATACGT TAAAATTGAT TCTAATTTTG TATGTCTTGA	4200
	TTGCTATAAG AATAACTTTA TTAATATCTA AAATTTAACA CTTAATGAAC TTGTTTCAAT	4260
5	GATATATTAG CACTATTTGT ATTTTGTGAT AACTAATATG TTTTGCATTT ATTTATAGTT	4320
	ATACTTCAAA TTACAAACTt CGCCATTTCA TATACCTTTT AATATCTATT TTGTTTTCGT	4380
10	CAACTACAGT TTTTATAATG ATACTGTATC TTCGATTTTT TTAGCAAAAA CAATTCTTCC	4440
	TGAAGATGTT TGCAATAAGC TGACTIONC TAAATTGACA TGACTIONC TAAGATTTTT	4500
	AGCATTATCA ACAACTACCA TCGTACCATC ATCTAGATAT CCTACTGCCT GACCAGGctC	4560
15	CTTACCCATT TTTGTCAGTA AAATATGCAG TTGATCACCT TGATGTACAT TAGGTTTGAT	4620
	TGCTTCTGAT AAATCATTAA CATTTAATGC TTTGATACCA TGTACATGAC AAACCTTTATT	4680
	TAGGTTGAAA TCTGTCGTTA TAATACTTGC ATGATATTGT TTTGCTAATT TTAATAACAT	4740
20	CGTATCAATA TCACTATGTG TTTTAGTTGG ATGTATAACC TTTGTAGGAT AGTCTAAATC	4800
	ATACAATTCA TTTAAAATAT CTAAGCCTCT TTTACCCCTT TCaCGTTTAA CACTGTCATT	4860
	TGAATCTGCA ACAATTTGTA ATTCATTAAT AACACCTTGT GGAATTAAAA TATTGCCATC	4920
25	GATAAAACCG CAACGAATGA CTTCTAAAT ACGACCATCA ATAATTGCGC TTGTGTCGAT	4980
	AATTTTTGGC GTAgcaCTTT TaGTATGTTG TGACATGGAA CGCGCTATAT TCTCAGGTAA	5040
30	AAACATTAAC ATTTTCATCTC GTTTTTTAAG GCCAAATTGG AAACCGAAAT AACATAGTAA	5100
	TATCGTAATT ATGACAGGAA TGAAATGATT AAAAATAGAG TTGCCAATTG ATTCTAATAT	5160
	AAACGACACC ATAACAGAAA TAAGTAATCC GATTATTAAA CCTATTGTTG CGAATAGTAT	5220
35	TTCAACAGCA CTTCTACGCA TAATAAAATG TTCTAAACCT TTTATAGCGT TAGTAACTCG	5280
	TCTAATAAAT ACACCAAAAA TTAAGAACAT AAAAATACTA CCGATAATGC CATCTACATA	5340
	GTGATTTTTT AAAAAGCTGG AGTTTTGTAA TCCAAGATCA TTTGCAATTT CAGGAATAAT	5400
40	AATTATTCTT AATGCGCTCC CAATAATTAA GTAAATAATA ATAACCATTA GTTTAACGAT	5460
	ATTCACACAA TGCTCTCCTT TCTTGATGTT TTATGAATGA AGAGCAAATG ACAATACTTC	5520
45	ATGTACAGTA GTTACACCTA TTACTIONTAT ACCTTCAGGA TATGTCCATC CGCCTATATT	5580
	ATTTTTAGGA ATAATTACAC GTTTGAAACC TAGTTTTGCA GCCTCTTGCA CGCGTTGTTC	5640
	TATCCGAGAT ACACGACGTA CCTCACCCGT TAAACCAACT TCTCCAATAT AGCAATCTAA	5700
50	TCCGTCGACA GCTTTATCTT TAAAGCTAGA TGCAAGTTGCT ACAATTACAC TTAAATCAAC	5760
	TGCTGGCTCC GTTAACTTTA CACCGCCAGC TACTTTGATA TAAGCATCTT GTTGTGTAA	5820
55	TAGATAATTT TCTTTCTTTT CCAAAACAGC CATCAACAAA CTTAATCGAT TATGATCAAT	5880

TATTAAAAGT GGTCTGGTTC CCTCCATGGT TGCAACAATT GTTGAACCTG GAACATTTGT 6000
 TGAACGTTCT TCTAAAAACA TTTCAGATGG ATTATTTACA CCTTTTAATC CACTTTGCTT 6060
 5 CATTTCGAAG ATTCCcATTT CATTTCGTTGA ACCAAAACGG TTTTTAACAG CTCGCAAAAT 6120
 TCGATATGCG TGGTGTTCAT CGCCTTCAAA ATAAAGCACA GTATCaACCA TGTGTTCTAG 6180
 CAATCTTGGG cCCAGCAATT TGACCTTCTT TCGTTACATG ACCCACTATA AAAGTTGCaA 6240
 10 TGTTCATTTG TTTAGCAATA TTCATTAAAC TTTGTGTACT TTCACGAACT TGTGAAACAG 6300
 AACCTGGCGC AGAGCTGATT TCAGGATGAT ATATTGTTTG AATCGAATCC ACTACTAATA 6360
 15 AATCAGGTTG TTCTTCTTTT ACTGTTTGAT AAATAACTTC AAGATCTGTT TCAGCTAATA 6420
 CTTGCAATTC ACTTGAATCT TCATCTAATC GCTCTGCACG TAATTTAGTC TGAATAAGCG 6480
 ATTCTTCTCC AGTAATATAT AGTACTTTTT TCTTTTGAGA TAACGATGCA CAAATTTGTA 6540
 20 AAAGTAACGT TGAATTACCA ATACCTGGAT CCCCACCAAT AAGTACTAAC GATCCGCTCA 6600
 CAATACCTCC ACCTAATACA CGGTTGAATT CTGCTGAATC TGTTAACACT CTCGGCGTTG 6660
 TTTTCATGTTT AATACTATTT AATTTTTGTA CTTTACCTGC TAATTCCTTG GTTTTAACTC 6720
 25 CATGTTTAGG ATTGGCTGCT TTTTCAACAA TTTCTCCAT TTGATTCCAA GCGCCACAAT 6780
 TAGGACATTT CCCCATCCAT TTAGGAGATT GATAACCACA AGCCATACAT TCAAAAATCA 6840
 CTTTTTCTT GGCCArAATT GCAcCTCCAC TTTCTT 6876
 30

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

CAACTCAAAC AGCAGAACAA CGTCGTGAGT TGATTAATGG TGTATTTACT GACATTAATC 60
 CCATACATTA AAAATATGAT GTACGTGTTA GCAGATAATA GACATATCTC ATTAATAGCT 120
 45 GACGTATTCA AGGCGTTCCA AAGCTTATAT AACGGACACT ACAATCAAGA TTTTGCAACA 180
 ATTGAGTCAA CATATGAATT GAGTCAAGAA GAGTTAGATA AGATTGTCAA ACTAGTAACT 240
 CAACAAACGA AGTTATCTAA AGTTATTGTA GATACAAAAA TTAATCCAGA TTTAATTGGT 300
 50 GGATTTAGAG TTAAAGTCGG CACAACGTGA TTAGATGGTA GTGTTAGAAA TGATCTTGTC 360
 CAATTACAAA GAAAATTTAG AAGAGTTAAT TAATTATAAA GAGGAGTGAC ATAGATGGCC 420

55

ATGTCCGTAA CTGATGTAGG TACTGTATTA CAAATTGGTG ATGGTATTGC ATTAATTCAC 540
 GGATTAAATG ACGTTATGGC TGGTGAGCTA GTAGAATTCC ATAACGGCGT ACTTGGTTTA 600
 5 GCCCAAAACC TTGAAGAGTC AAACGTGGGT GTGGTTATTT TAGGACCATA CACAGGTATT 660
 ACTGAAGGTG ACGAAGTTAA ACGTACTGGT CGTATCATGG AAGTACCAGT AGGTGAAGAA 720
 10 CTAATCGGAA GAGTTGTTAA TCCATTAGGA CAACCTATTG ATGGACAAGG ACCGATTAAAC 780
 ACAACTAAAA CACGTCCaGT AGAGAAAAAA GCTACTGGTG TAATGGATCg TAAATCAGTA 840
 GATGAGCCAT TACAAACAGG TATCaAAGCA ATTGATGCTT TAGTACCAAT TGGTAGAGGT 900
 15 CAACGTGAGT TAATCATCGG TGACCGTCAA ACAGGTAAAA CAACAATTGC AATTGACACA 960
 ATTTTGAACC AAAAAGATCA AGGTACGATT TGTATCTATG TTGCTATTGG TCAAAAAGAT 1020
 TCAACAGTAA GAGCAAATGT TGAAAAGTTA AGACAAGCAG GCGCTTTAGA CTACACTATT 1080
 20 GTTGTAGCAG CATCAGCTTC TGAACCTTCT CCATTATTAT ATATTGCACC ATATTCAGGT 1140
 GTAACATGG GTGAAGAATT CATGTTTAAC GGTAAACATG TTTAATCGT TTA 1193

(2) INFORMATION FOR SEQ ID NO: 188:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5549 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

TGCTAAGAAG TCAAAATAAA CTAACtATnA AACATCTAGT ACGATTATTA AAGTGACAGA 60
 TnATAAAATT GAATTATtNa GAGAAGGAGA TATAAAGTTT GAAGAAATAA AAGAAAGACT 120
 AGGTACAGGT ATTATTTATG AATAAGTTAA TACTTGGGAT TTATTTATAC CGAATTTTTT 180
 40 CACGAGCATA CTTTTATTTA CCGTTTTIAT TAATTTACTT TTTGATTCAA GGTATTCCA 240
 TAATACAATT AGAAATATTA ATGGCGTCTT ATGGCATTGC AGCATTTTTA TTCTCTCTAT 300
 ACAAGAGAA GTGTTTTTAAA ATTTGTAACT TAAAAGATTC TAATAAATTA GTTGTTAGTG 360
 45 AAATATTCAA AATCATCGGT TTATTGTTGT TATTATATCA AAATCAATAT TTAATTTTAG 420
 TAGTGGCACA AATATTATTA GGGTTAAGTT ACTCAATGAT GGCGGGTGTT GATACCGCAA 480
 50 TAATTAAAAG AAATATAACA AATGAGAAAT ACGTACAAAA TAAGTCAAAT AGCTATATGT 540
 TCCTATCATT ATTAATTTCA GGGATTATAG GTAGTTATCT TTATGGAATA AATATTAAAT 600
 GGCCTATAAT AATGACTGGT ATATTTTCAA TTCTAACAAAT TATAATTATT CGATGCACAT 660

TACCAGAAGA GAAGTTTTGG ATATTGCATT ATTCTTTTTT AAGAGCGTTA ATATTAGGAT 780
 TTTTATAGG ATTTATTCCA ATTAATATAT ATAATGATTT AAAACTGAAT AATTTACAAT 840
 5 TTATTTTCAGT ATTAACCTGT TACACAGTTA TGGGTTTTGT ATCTTCACGT TATTTAACTA 900
 AATACTTGAA TTATAAGTTT GTGTCAGAAA TTTGTTTAGT AATATTTTTA ATAATATATA 960
 10 CATATCAAAG TTTCATAGCA GTTACTATTT CTATGATATT TTTAGGTATT TCTTCAGGGT 1020
 TAACTCGTCC ACAAACTATA AATAAACTTT CTAGCAGTAG TAACTTAAGA GTGATGCTTA 1080
 ATTATGCAGA AACGTTATAT TTTATTTTTA ATATCGCATT TTTACTTATG GGTGGTTACT 1140
 15 TATATACAAT AGGAACTATT CAATACTTAA TATTATTTAT TTCGTTATTA ATTTTTATAT 1200
 ATTTAATAAT AATATTTyAT TTTACAAGGA GAGAGCAACA TGAAAATAAA AACTGAATTT 1260
 AAAGGGAACA ATATACCATA TGAATACGCA GCAGGTGCAG ATGTGAGTGA TTCTATTAAAC 1320
 20 GGGAAATCCAA TTAAGTCATT TCCATTTGAA GTAATTGAAT TACCGGAAGG gACTAAATAT 1380
 CTTGCTTGGT CTTTAATTGA CTATGATGCA ATTCTGTAT GTGGCTTTGC TTGGATTCAT 1440
 TGGAGTGTAG CTAATGTAAG TGTTAGTGGC AATTCAATTT CTATAAAAGC AGATTTATCA 1500
 25 AGAACAAAGG GCGACTATGT ACAAGGTAAA AATAGCTTTA CTAGTGGGTT GTTGGCTGAA 1560
 GATTTTTTCAG AAATAGAAAA TCACTATGTA GGACCTACAC CACCTGATCA AGATCATCAA 1620
 TATGAATTAA CAGTTTATGC GTTAGATCAT TCTTTAAATT TGAAGAATGG GTTCTACTTG 1680
 30 AATGAATTTT TAAAGAAGT AAATCAACAT AAAATTGATC AAACAAGTAT TAACCTTATA 1740
 GGAAGAAAAA TTTAATACTA AATATCTCAT CAATATAAAA TTGTTCAATT AAAAGTACAA 1800
 35 AGAAACAAAG GTTTTAATTT ATATATTAGG TACGGCGTTC GCTATAATGC AAAGAAGTAA 1860
 TTAAATTTAA GAAATGTAAA CTTAGTTATT GTAATGTGAA TTTATTTGAA AAAATAGAAA 1920
 GTATTAACAA TTATAGCTTT TACATTAATT AAAATTTATT TTTAAAAACA AGTAAACAAT 1980
 40 TTACATACTT ATAATTTTTG AAAATTTTCA ATTTGTGTTA TATTGATTTT GTAAGATACT 2040
 TTAACTCACA AAGGAGAGAG AGTATATGAA ATTAAATCA TTTATAACTG TAACTTTGGC 2100
 ACTGGGCATG ATCGCAACGA CTGGCGCTAC TGTGGCAGGT AATGAGGTAT CTGCAGCAGA 2160
 45 AAAGGACAAA CTACCGGCAA CTCAAAAAGC TAAAGAAATG CAAAATGTTC CATATACAAT 2220
 TGCAGTAGAT GGCATTATGG CTTTCAATCA ATCTTACTTA AATTTACCAA AAGATAGCCA 2280
 ATTATCATAT TTAGATTTAG GAAATAAAGT TAAAGCTTTG TTATATGATG AACGCGGTGT 2340
 50 AACACCTGAG AAGATTGCAA ATGCAAAATC TGCCGTTTAC ACGATTACTT GGAAAGATGG 2400
 TAGTAAAAAA GAAGTGGATC TTAAGAAAGA TAGCTACACA GCAAACCTGT TTGATTCAAA 2460

55

	CAACATGAAG	CATTTAATTT	TACAGTGATG	ATTATAAAAT	AATTGCCTTG	ATACAAAGAT	2580
	TACTCGTAAA	TGACATCTTT	GTATTAAGGC	TTTTTCTAAA	TTTAAAAGTG	ATGGGTTAGA	2640
5	GGTCATTGAG	CTTTAAAATA	TTCAAAATAC	AAAACATTAA	TGGCCAAAAA	TAAAAGCCGC	2700
	CTTTATCTGG	GCAGCTTCAA	TAATAAGAAA	GACATATTTT	ATTTTATACT	AAATAGTTAT	2760
	TGTGATGAAT	CTTTCGGCGG	TTTAATTACT	GCAGCAAAAA	TTGCTGTGAA	AATCGTGAAC	2820
10	AATACTGCCA	TGATAATTGG	ATTCACTACA	TTTAAGCTGT	CTCCACCTAC	TAGGCTATTA	2880
	AGTACAAAGT	TAACCATTTG	CATTAATAAT	AATGCCCAAA	AGAATGTTAC	GAGGTGTTTC	2940
	ATGTCATTCT	ACCTCCACTT	TAATTATATA	TATTTTATTT	TAAGTGAAAG	TTAGAAATTT	3000
15	GTATAGTAAC	ATCTCATATA	TTTTGACCAT	ATTATACAGT	TTAAATAAAT	GATTTTATCT	3060
	GAATGGCTAT	TCTAAATTAA	GCGCATTAAA	ACCAATTTC	TACTGAAATT	TGACGATAAT	3120
20	AAAGCATTA	AATTTTATTA	ACTAGTCAAT	ATTCTACCT	CTGACTTGAG	TTTAAAAAGT	3180
	AATCTATGTT	AAATTAATAC	CTGGTATTAA	AAATTTTATT	AAGAAGGTGT	TCAACTATGA	3240
	ACGTGGGTAT	TAAAGGTTTT	GGTGCATATG	CGCCAGAAAA	GATTATTGAC	AATGCCTATT	3300
25	TTGAGCAATT	TTTAGATACA	TCTGATGAAT	GGATTTCTAA	GATGACTGGA	ATTAAAGAAA	3360
	GACATTGGGC	AGATGATGAT	CAAGATACTT	CAGATTTAGC	ATATGAAGCA	AGTTTAAAAG	3420
	CAATCGCTGA	CGCTGGTATT	CAGCCCGAAG	ATATAGATAT	GATAATTGTT	GCCACAGCAa	3480
30	CTGGaGATAT	GCCATTTC	CAACTGTCG	CAATATGTTG	CAAGACGTTT	GGGACGGGCA	3540
	AAGTTGCCTC	TATGGATCAA	CTTGCAGCAT	GTTCTGGATT	TATGTATTCA	ATGATTACAG	3600
	CTAAACAATA	TGTTCAATCT	GGAGATTATC	ATAACATTTT	AGTTGTCGGT	GCAGATAAAT	3660
35	TATCTAAAAT	AACAGATTTA	ACTGACCGTT	CTACTGCAGT	TCTATTTGGA	GATGGTGCAG	3720
	GTGCGGTTAT	CATCGGTGAA	GTTTCAGATG	GCAGAGGTAT	TATAAGTTAT	GAAATGGGTT	3780
40	CTGATGGCAC	AGGTGGTAAA	CATTTATATT	TAGATAAAGA	TACTGGTAAA	CTGAAAATGA	3840
	ATGGTCGAGA	AGTATTTAAA	TTTGCTGTTA	GAATTATGGG	TGATGCATCA	ACACGTGTAG	3900
	TTGAAAAAGC	GAATTTAACA	TCAGATGATA	TAGATTTATT	TATTCCTCAT	CAAGCTAATA	3960
45	TTAGAATTAT	GGAATCAGCT	AGAGAACGCT	TAGGTATTTT	AAAAGACAAA	ATGAGTGTTC	4020
	CTGTAAATAA	ATATGGAAAT	ACTTCAGCTG	CGTCAATACC	TTTAAGTATC	GATCAAGAAT	4080
	TAAAAAATGG	TAAAATCAAA	GATGATGATA	CAATTGTTCT	TGTCGGATTC	GGTGGCGGCC	4140
50	TAACTTGGGG	CGCAATGACA	ATAAAATGGG	GAAAATAGGA	GGATAACGAA	TGAGTCAAAA	4200
	TAAAAGAGTA	GTTATTACAG	GTATGGGAGC	CCTTTCTCCA	ATCGGTAATG	ATGTCAAAAC	4260
55							

TGAACCTTAT AGCGTTCACT TAGCAGGAGA ACTTAAAAAC TTAAATATTG AAGATCATAT 4380
 CGACAAAAAA GAAGCGCGTC GTATGGATAG ATTTACTCAA TATGCAATTG TAGCAGCTAG 4440
 5 AGAGGCTGTT AAAGATGCGC AATTAGATAT CAATGAAAAT ACTGCAGATC GAATCGGTGT 4500
 ATGGATTGGT TCTGGTATCG GTGGTATGGA AACATTTGAA ATTGCACATA AACAATTAAT 4560
 10 GGATAAAGGC CCAAGACGTG TGAGTCCATT TTTCGTACCA ATGTTAATTC CTGATATGGC 4620
 AACTGGGCAA GTATCAATTG ACTTAGGTGC AAAAGGACCA AATGGTGCAA CAGTTACAGC 4680
 ATGTGCAACA GGTACAAATT CAATCGGAGA AGCATTTAAA ATTGTGCAAC GCGGTGATGC 4740
 15 AGATGCAATG ATTACTGGTG GTACAGAAGC ACCAATTACT CATATGGCAA TTGCTGGTTT 4800
 CAGTGCAAGT CGAGCGCTTT CTACAAATGA TGACATTGAA ACAGCATGTC GTCCATTCCA 4860
 AGAAGGTAGA GATGGTTTTG TTATGGGTGA AGGTGCTGGT ATTTTAGTAA TTGAATCTTT 4920
 20 AGAATCAGCA CAAGCTCGAG GTGCCAATAT TTATGCTGAG ATAGTTGGCT ATGGTACTAC 4980
 AGGTGATGCT TATCATATTA CAGCGCCAGC TCCAGAAGGT GAAGGTGGTT CTAGAGCAAT 5040
 GCAAGCAGCT ATGGATGATG CTGGTATTGA ACCTAAAGAT GTACAATACT TAAATGCCCA 5100
 25 TGGTACAAGT ACTCCTGTTG GTGACTTAAA TGAAGTAAA GCTATTAAAA ATACATTTGG 5160
 TGAAGCAGCT AAACACTTAA AAGTTAGCTC AACAAAATCA ATGACTGGTC ACTTACTTGG 5220
 TGCAACAGGT GGAATTGAAG CAATCTTCTC AGCGCTTCA ATTAAAGACT CTAAAGTCGC 5280
 30 ACCGACAATT CATGCGGTAA CACCAGATCC AGAATGTGAT TTGGATATTG TTCCAAATGA 5340
 AGCGCAAGAC CTTGATATTA CTTATGCAAT GAGTAATAGC TTAGGATTCTG GTGGACATAA 5400
 CGCAGTATTA GTATTCAAGA AATTTGAAGC ATAACTATAA nAATCTTCAG TAACGTTGTT 5460
 35 TTAGTTACTG AAGATTTTTT CaGTTTCTTT ATACTAAGAT GAGCGACaCa CAATCGTCAT 5520
 AATAAAATAT GAATATTTAT TAATAATAA 5549

40 (2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4832 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

AGATTATAGT AAGATTGATA GTTTGGCGAC TGaAGcCGa GaAAAAATTAT CAGaAGTAAA 60
 mCCTTTAAAT ATTGCACAAG CTTCTAGAAT ATCAGGGGTA AATCCAGCAG ACATATCTAT 120

55

	TGGTTAGCAG	AACAATTAAA	AGAACATAAT	ATTCAATTAA	CTGAGACTCA	AAAACAACAG	240
	TTTCAAACAT	ATTATCGTTT	ACTTGTTGAA	TGGAATGAAA	AGATGAATTT	GACAAGTATT	300
5	ACAGATGAAC	ACGATGTATA	TTTGAAACAT	TTTTATGATT	CCATTGCACC	TAGTTTTTAT	360
	TTTGATTTTA	ATCAGCCTAT	AAGTATATGT	GATGTAGGCG	CTGGAGCTGG	TTTTCCAAGT	420
10	ATTCCGTTAA	AAATAATGTT	TCCGCAGTTA	AAAGTGACGA	TTGTTGATTC	ATTAAATAAG	480
	CGTATTCAAT	TTTTAAACCA	TTTAGCGTCA	GAATTACAAT	TACAGGATGT	CAGCTTTATA	540
	CACGATAGAG	CAGAAACATT	TGTAAGGGT	GTCTACAGGG	AGTCTTATGA	TGTTGTTACT	600
15	GCAAGAGCAG	TAGCTAGATT	ATCCGTGTTA	AGTGAATTGT	GTTTACCGCT	AGTTAAAAAA	660
	GGTGGACAGT	TTGTTGCATT	AAAATCTTCA	AAAGGTGAAG	AAGAATTAGA	AGAAGCAAAA	720
	TTTGCAATTA	GTGTGTTAGG	TGTAATGTT	ACAGAAACAC	ATACCTTTGA	ATTGCCAGAA	780
20	GATGCTGGAG	AGCGCCAGAT	GTTCAATTATT	GATAAAAAAA	GACAGACGCC	GAAAAAGTAT	840
	CCAAGAAAAC	CAGGGACGCC	TAATAAGACT	CCTTTACTTG	AAAAATAATG	CATAATCCTT	900
	TACAACCTAAC	ATAAAAGGAG	CGAATGGATA	ATGAAAAAAC	CTTTTTCAAA	ATTATTTGGT	960
25	TTGAAAAACA	AAGATGACAT	CATTGGACAT	ATTGAAGAAG	ATCGCAATAG	TAATGTTGAA	1020
	TCCATTCAAA	TTGAACGTAT	CGTTCCCAAC	CGTTATCAAC	CAAGACAGGT	GTTTGAACCA	1080
30	AATAAAATTA	AAGAACTTGC	TGAATCAATA	CATGAACATG	GTTTACTACA	ACCTATTGTT	1140
	GTAAGACCGA	TTGAAGAAGA	TATGTTTGAA	ATTATTGCTG	GAGAGCGCCG	ATTAGAGCA	1200
	ATACAATCAC	TAAATTTACC	TCAAGCAGAC	GTTATTATTC	GTGATATGGA	TGATGAAGAG	1260
35	ACGGCTGTTG	TTGCATTAAT	TGAGAATATT	CAAAGAGAAA	ATTGTCTGT	TGTTGAAGAA	1320
	GCGGAAGCCT	ATAAGAAATT	ATTGGAAATT	GGTGATACAA	CGCAAAGTGA	ATTGGCAAAA	1380
	AGTTTAGGTA	AAAGTCAAAG	CTTTATTGCA	AATAAGTTGC	GTTTATTGAA	GTTGGCGCCG	1440
40	AAAGTACTAC	TTCGCTTAAG	AGAAGGTAAA	ATTACTGAAC	GTCATGCGAG	AgcGGtATTA	1500
	TCATTGTCTG	ATAGCGAACA	AGAAGCGTTG	ATTGAGCAAG	TCATTGCACA	AAAGCTAAAT	1560
45	GTGAACAGAC	TGAAGATAGA	GTACGCCAAA	AAACGGGGCC	CGAAAAAGTC	AAAGCACAAA	1620
	ACCTTCGCTT	TGCACAAGAT	GTCACTCAAG	CACGAGATGA	GGTAGGCAAA	AGTATCCAAG	1680
	CGATTCAACA	AACAGGATTA	CATGTTGAGC	ATAAAGACAA	AGATCATGAA	GATTATTATG	1740
50	AAATAAAAAT	TCGAATATAT	AAACGTTaGT	AGTAGGATGT	CGTATACATG	ATGACTAACA	1800
	CATAAAAGAC	AAAGCTAAGA	TCATAACAGC	TTTGTCTTTT	TTTTTTGTTT	TACGTGAAAC	1860
	ATAAAAATTT	ATATTTATAT	GTTGATCAGG	CTGGTACATA	AATCAATGTT	CTATGCTCTA	1920

55

	TTCTAGTCAA CCTTGCTGGG GTGGGACGAC GAAATAAATT TTGCGAAAAT ATCATTTCTG	2040
	TCCCACCTCCC TAATTTGAGC TGGATATACT TTCATTTGAA CCCTTTATTG CTAGTTTATG	2100
5	AAAGTATCAT GAAAGCTTTA TGAACATCGC TTGAGTTGCC TTTACAGTAG AAAATTTAAG	2160
	TTTTACACTT TGTGTGAATG ATACGTTTGT TATTGAATTA ATTATAGAAA GGTACGTTGA	2220
	AGATGTTTTT AATTGGAAGT GCAATTCTTC ATTTTGTCTAT TGGTGGTATC GCTGTTGCAT	2280
10	TAGCTTCAAT TATTGCTGAT AAGGTAGGTG GTAAGTTAGG AGGTATTATA GCTACTATGC	2340
	CGGCAGTCTT TCTTGCGGCT ATTATCGCAT TAGCTTTAGA TCATCGTGGT ACGCAATTAG	2400
	TGGAGATGTC GATGAATCTT AGTACTGGAG CAATTGTCCG TATTCTGTCT TGTATATTAA	2460
15	CTGTATTTTT GACATCTCTC TACATTAAGC ATAAAGGTTA TCGGAAAGGC GCAATATTCA	2520
	CAGTTGTTTG TTGGTTTGTG ATTTCCCTCG CAATATTCAG TATTAGACAT TTATAGTTTG	2580
20	GAAAATGCGT GATAATTAGT TGTATTCAGT TATTAAGTAA TAAATTATTG GAGGCAGAAC	2640
	ATCATGAAAT TAACATTAAT GAAATTTTTT GTGGGGGGAT TTGCAGTATT ATTAAGTTAT	2700
	ATTGTATCTG TAACACTACC TTGGAAAGAA TTTGGCGGTA TATTTGCaAC GTTCCGGCA	2760
25	GTATTTTTAG TGTCTATGTT TATTACAGGT ATGCAATATG GTGATAAAGT CGCTGTGCAT	2820
	GTAAGTCGTG GCGCAGTGTT TGGTATGACA GGGGTATTAG TTTGTATTTT AGTTACATGG	2880
	ATGATGTTAC ATATGACGCA CATGTGGTTG ATTAGCATTG TTGTTGGTTT CCTAAGCTGG	2940
30	TTTCATCAGTG CAGTATGTAT TTTTGAAGCG GTAGAATTTA TAGCACAAAA AAGATTAGAA	3000
	AAGCATAGTT GGAAAGCTGG AAAATCGAAT AGTAAATAGT GTGAACGTAA TCTCTTAACT	3060
35	AGGACTAACT TTGCAAGCAT TGAATAGCAT GGAAAAGTTG CATCATTAAAT AAGTGAAATT	3120
	CAAGTTGGCA TTGAGAAAAT TACAAGCGCG TAATCATACa GGTCTGTCTT AAGGGAGTCT	3180
	TCGAACCCCG ATGTTGTCGT ATGTCAAAAC ATTTAGTCAA TCATAAAGGT GACTTGATTT	3240
40	AACTTTATCT GATAGTCTGA TTGTAATGAT TGTACTAATT GACTGGAGGC GTATGTAATT	3300
	GAATCTGAGT AAACAAATTA AAAAGTATAG GGAACGAGAT GGTATTTCAC AAGAATATCT	3360
	TGCTGAAAAG TTATATGTAT CTAGGCAGAG TATTTCTAAT TGGGAAAATG ACAAAGCTT	3420
45	ACCAGACATA CATAACTTAT TAATGAYGTG TGAATTGTTT AATGTAACCT TAGATGATTT	3480
	AGTAAAAGGG ACCATTCCAT TTGTACCTGA TATTAAAGCG CAACGAAGTC TTAACCTATG	3540
	GACATATGTG ATGCTTATTT TCATGACATT AGCTGCAATT TTAATGGGAC CTTTAGTTGT	3600
50	TTATTGGAAT TGGACTTGGG GTGTAACGGT GGCAATCATT TTGGGAATAG GTTTTTATGC	3660
	ATCTATGAAA ATAGAAGATT TAAAAAAGT GCATAAAATG GACAACTACG ATCGAATTGT	3720

55

5 GACAAATGCG CTTTCTATTA TATCAGTAAT TGGTATACTC AGCCTCATAA TTTTCCTTAG 3840
 TGTGTATTTG GCAAATAAGT TTTTATAAAT CATCGTGGTA TCGTCTCATA TTATTTATAT 3900
 10 TATCCAAAAT AGCATAAAAA AATACCAACA AGATTTAGAA CCTTGTTGGT AATCAAAGCG 3960
 aTTCATTTAT AATGAGTCGT TTTATGTTGT AAGATTAAAC AGTTTGTACG TTAAGTGCCT 4020
 GGTCTCCACG TTGACCTTCA GTGATTTCGA AAGTAACTTT TTGACCTTCT TCTAAAGTTT 4080
 15 TGTAGCCATC GCTAGCGATA CCTGAGAAAT GTACGAATAC GTCTCCGCCA TTTTCTTGTT 4140
 CGATGAAACC AAAACCTTTT TCTGCTTTAA ACCATTTWAC TGTACCGTTA TTCATATWGA 4200
 AwACCTCCGT gTGCTTTTGC ACTTAATATT TGTAACAAAT TCATAACTAA AAAAGAGGAT 4260
 ATTCTAAACA AATACACTAC AATTTAATTC ACGAGCTTTT ATTACGTAAG ACCAACTATA 4320
 CGCTCATATT GGCATAATGT ACAGTGTTTT TTGAAAATAA ATTAAAAAAG ATTTTAAAAA 4380
 20 ACCTTAGAAA CGTTGATTTA AAGGGGTTTA TAAAAATWAw AAAATTGTAG TCTTTTATGG 4440
 TGTTTGCTAG TTTTCAAAGT GACATATCGT TTAAACATGA TGATTTTATA AGCAATCCAT 4500
 AAAAAACAAG CAGCGATAAA CGCTACTTGT TGATATTAAA ATCTGACTTG AAAGGTCATA 4560
 25 GCAATGTTCT ATACCGATGG AATGTGCTTA CTTGCCTTTT TCTTCACGAC GTTTTAAATA 4620
 ATAAGAGCCA CCTAATAAAC CAGCTGGAAT GCCTATCATT GGTGTTGTGA ATGAGCTTAA 4680
 TACAATAACA AGTATTGTTA AAGCAATGAC GTTATACCAA GTTACAGTCA AATTTTTCAA 4740
 30 ATCCTCATAT GATTGTTTTA CTAATTCTCT AAATTTTCATG ATTCAATCTC TCCTTTTTTA 4800
 TAAATCTTTA GATTGTCAA TTAAGCTGGA CA 4832

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5727 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

45 CAAAGCTGTT CAAAAGGCTT ATAATTTAAA TTAGATAAC ATACGTACAA TGGAACCTAA 60
 GTTGAGATAT CAAGCGATCA ATAAAGGTAA TATTAATTTA ATAGATGCAT ATTCAACTGA 120
 CGCTGAATTA AAACAATATG ATATGGTTGT GTTAAAGAT GATAAGCACG TATTTCCACC 180
 50 ATATCAAGGA GCACCATTAT TTAAAGAAAG CTTTTTAAAG AAACATCCAG AAATTAAGAA 240
 ACCGTTAAC AACTAGAAA ACAAATATC TGATGAAGAT ATGCAAATGA TGAAGTATAA 300

	GTTAATCAAA TAACGACCAA CGCCACATAA GATGCGTAAC ACCAAATTAT ATCTTATGTG	420
	GCGTTGTTAT ATTAAATCT ATAATTATGT TCAATTTAAA CATGCAATAA TGATTAAAAA	480
5	ATATGACATG TTAACACAA TGTAAGCTAT TATGATGTGA AAATAGTAGC ATTGCATTTT	540
	AGAAACATAG AGCGATATAA TGAATATAAG TTTTTTGAAA TTTCAGTTAA TTCTAAGGAG	600
	GTTGTTTTTA TTATGAAAGA ACAACTTAAT CAACTATCAG CATATCAGCC TGGTTTATCT	660
10	CCAAGGgCAT TGAAAGAAAA GTATGGCATT GAAGGAGATT TATATAAACT TGCATCAAAT	720
	GAAAAATTTGT ATGGACCATC GCCTAAAGTT AAAGAAGCGA TATCAGCACA CTTAGATGAG	780
15	TTATATTATT ATCCTGAAAC AGGATCACCG ACATTAAAAG CGGCGATTAG TAAACATTTA	840
	AATGTAGATC AATCACGCAT TTTATTTGGT GCGGGATTAG ATGAAGTTAT ATTAATGATT	900
	TCTAGAGCTG TATTAACGCC AGGGGATACT ATTGTTACAA GTGAAGCGAC ATTCGGTCAA	960
20	TATTATCACA ATGCGATTGT TGAATCAGCT AATGTGATAC AAGTACCTTT AAAAGATGGT	1020
	GGCTTCGATT TAGAAGGTAT TTTAAAAGAA GTTAATGAAG ATACGTCATT GGTATGGTTA	1080
	TGTAATCCAA ATAATCCTAC AGGTACATAT TTTAATCATG AGAGCTTAGA TTCGTTTTTA	1140
25	TCTCAAGTAC CTCCACATGT ACCAGTAATT ATAGATGAAG CTTATTTTGA ATTTGTGACA	1200
	GCAGAGGACT ACCCGGATAC ACTTGCTTTG CAACAAAAAT ATGACAATGC TTTCTTATTA	1260
	CGTACATTTT CAAAGGCGTA TGGATTAGCG GGTTTACGTG TAGGATATGT GGTAGCAAGT	1320
30	GAACATGCGA TTGAAAAATG GAACATCATT AGACCACCAT TTAATGTGAC ACGTATATCT	1380
	GAATACGCAG CAGTTGCAGC ACTTGAAGAT CAACAATATT TAAAAGAGGT AACACATAAA	1440
	AATAGTGTG AACGCGAAAG ATTTTATCAA TTACCTCAA GTGAGTATTT CTTGCCAAGT	1500
35	CAAACGAATT TTATATTTGT AAAAACmAA GGGGTAAATG AACTTTATGA AGCACTTTTA	1560
	AATGtAGGGT GTATTACGCG ACCATTTCCA ACTGGTGTGA GAATTACAAT TGGTTTTAAA	1620
	GAACAAAATG ATAAAATGTT AGAAGTTTTA TCAAACTTA AATACGAATA GTAAGTGGGG	1680
40	AGTGGGACAG AAATGATATT TTCGCAAAAT TTATTTcGtC GTCCCACCCC AACTTGcATT	1740
	GTCTGTAGAA ATTGGGAATC CAATTTcCT TTGTTGGGGC CCCGCCGGCA AGGTTGACTA	1800
45	GAATTGAAAA AAGCTTGTTA CAAGCGCATT TTCGTTcAGT CAACTACTGC CAATATAACT	1860
	TTGTAGAGCA TTGAACATTG ATTTATGTCT CAAGCTCAAT GCAGTGTGAA TGATGAGGTG	1920
	AGAGTATTCA GTGTAAAAAG CAACAATAGA TGATATTGTT TTGTATCAAT TGCTTTTTTG	1980
50	CTATACTGAA TCAATACTGA TATTTTCAGG AGAAGATTAA AATGACCCGT AAATCAATCG	2040
	CGATTGATAT GGATGAAGTA TTGGCAGATA CATTAGGAGA AATCATTGAT GCTGTCAATT	2100

55

	TTCCTGAACA TGATGGATTA ATTACAGAAG TATTGAGAGA ACCAGGCTTC TTCAGACATC	2220
	TTAAAGTGAT GCCGTATGCA CAAGAAGTTG TGA AAAAATT AACTGAACAT TATGATGTAT	2280
5	ATATTGCTAC AGCAGCAATG GATGTACCAA CATCATTTAG TGATAAATAT GAATGGTTAC	2340
	TAGAGTTCTT TCCATTTTTA GATCCTCAGC ATTTTGT TTTT TTGTGGTAGA AAAAAATCG	2400
10	TTAAAGCTGA TTATTTAATA GATGACAATC CTAGACAGCT TGAAATTTTT ACTGGTACAC	2460
	CGATTATGTT TACAGCAGTG CATAATATTA ATGATGATCG ATTTGAACGC GTAAATAGCT	2520
	GGAAAGATGT AGAACAGTAT TTTT TAGATA ATATTGAGAA ATAAATATA TCACTTGAAA	2580
15	AATTTTCATGT AGAAAAGATG ATGGATAGGC TATAAAGTAA TTGTGACTGA GATGAAC TTT	2640
	TATGTCTTAG ACACTACAAC ACTATATTGG CAGTAGTTGA CTGCGGGGCC CCAACATAGA	2700
	GAAATTGGAT TCCCAATTTT TACAGACAAT GCAAGTTGGG GTGGsCCCCA ACATAAAGAA	2760
20	ATACTTTTTT TTTAGAAATT AGTATTTCTT ATGCATGAGT GTAACTCATG CATT CATATT	2820
	TTTAAGTACA CATTAGCTGT GACTAATGAT AAAGAATCGC TACATAATCA ATCATTAGTC	2880
	GTTCTTTATC ATTTCCGTCC CGCTCTCAAT AAATGTTAGT CTATCTTATT ATTATAAATC	2940
25	GGATGAATGT GTTAATCTAT GGCAGATTAC ACGTCATCCG ATTTTTTATA GAATTTGAAA	3000
	AAGACGCATA AACCACTATG ATTTAAAATA CAACATCAAT CATTTTAGTG gCATGCGCCA	3060
	AAATTATATG TCTGTTTTTG AAACAGGGTA ATAGCTTAAA GCTAATAAAA ACGAATATAA	3120
30	GGTGCGTTGA ATCTTATGAT TACACTCCAA ACCTAATATA ATATCGGGTT AAGATCATTC	3180
	CGGATGCTTA CAAATCATTG ACAGTAAGTA ACTGAATGGC ATTTGGTATA ACCTCAATAT	3240
35	CAATAGGTGT TTCTAATGAA ATTTGCCCAT CAATATCAAC TTTCATTGCT GGATCTGTTG	3300
	TAAGTGAAAT CTTTTTACCA GGTATATGCT CAATACCTTG AGTAATTTCA TTCCaATTCA	3360
	TGCTATCAG CTTTTTAAAA ATATCATTTA AAATACTGAA ACTTTGTTCA TTAAAAATGA	3420
40	AAGTGTT CAG TTCACCATCT TGAGGAGACA AATCAGTCaA TGGTATACGA CTACCACCAA	3480
	TGAATGGACC ATTTGCTGTT AGTATCATGG TCGTTTCGCC AGAATATGTC TTATCATCTA	3540
	TTGATAATTG ATAATTAAAT TGTGTTGGAT TTAGCAGTGT TTTGACAGTT GATCCAATAT	3600
45	AACTCAATTT ACCAAATATA TCTTTTGAAC CATCTTGTAC GTTTTCAGCG TTTTGAACAA	3660
	TGAGACCTAA GCCAACAAAG TTGAGTGCAT ATTGATTATT TATTTTAATT ACATCGTATG	3720
	TACCAACTTG TGCAGAAATC ATTTGTTTAC TAGCTTGT TTT ATGATTAGGT GCTATATTTA	3780
50	GCGTTTTTGT AAAATCATTAA AAAGTACCGC CTGGTAAAT GCCAATAGGG AGTTGAAGGT	3840
	CATGTGTCAT AACACCGTTT ATAAGTTCGT TAACCGTGCC ATCACCGCCA AGAATAAATA	3900

55

	CACCTTCGTT TCACTCAAT TGAATAGAAA GATGCTTACA AATTGAACTT AATGCTGTTG	4020
	TAACTTCCCC AATACCTTGA TTAATATTTT TTAATCCACT GTGTTTCATGG TAAAAGAGGA	4080
5	CACCATGTGT ATATTTATTT TCCATAGTTT AGCCTACTTT CTAAAAATTG GTTCATTAAA	4140
	TATATATACC CACTTTTAAT TGTTAATACC AAAAATATGT TTTTAAATAG AGAAAATGGT	4200
	AATAAATGAA ATTGATTTCT ATAGAGTGGG ACGAGAAAAT ATAGTTATAG CTGTCTATAA	4260
10	TGAGCATATT AAGTTTTTAT TTATACTGAT ATCTTGAATT TAATTAATAG AAACCTATAA	4320
	AAAAACAGTA AGCCATTTAA ATGACTTACT GTTTTTTGAA TTAGGCCAAC AATATTAACG	4380
	TATACCTTTC ATCGCTTTGA TGATTAAAGG TGAGAATGCT AATACAATTG TTGTAACAAT	4440
15	AATTGCAACA ACACCTAGGA AAATAAAGTA ATTTGTTTGA CCTAGTGGTT CTATTAACCT	4500
	AACTAAAGTA CCATTGATTG CTTGTGCAGA AGCGTTAGTT AAGTACCAA TACTCATCAT	4560
20	TTGGGCATTA AATGCTTTAG GTGCTAACTT AACAGCAGCA CTATTACCCG TTGGTGATAA	4620
	GCATAGCTCA CCGATAACAC AAATAATGTA CGATAAAATA ACCCAGTTAA CTGAAAAGTT	4680
	TGATGAACCT GATGCATAAC CTACAATACC AATTAGTATG TATGACGCAC CTGCTAAGAA	4740
25	CGTACCAATT GCAAATTTTA CTGGCAGGCT AGGTTGTTTA GTTCCAAGCT TTTGCCATAA	4800
	AAGTGAAATA ATTGGAGCTA GTAATAAAAT AAATAATGGG TTAATTGATT GGAAGATCGC	4860
	TTCACCAAAG TTTGTTTCC AACCAAATAA GTTTAATTTT ATATCTGAAT GTTCAATTCC	4920
30	ATATATGTTT AATACATTAG ACCCTTGTTT TTGAATAGCC CAGAACACCA TTCCAAGAAT	4980
	AAATAATGGA ATAAATGCTT TAACACGAGA ACGTTCAGTA TCAGTGACAT CTTTACTTCT	5040
	AATAATTAAA GTGAAGTAAA TGAAATGGTAA TGCAATACCT AATACTAAAA CAGTATTACT	5100
35	AACTAAGTTA AATGATAATG AGTTAGTTAA TGCACCAATA ACGATAATTA ATACAATTGC	5160
	TAAACAACA CTTCCGATAA TAAGACCATA CTTTCTCTT TCAGCTGGTG TCAATGGGTT	5220
40	AGTAGGTTTC ATACCAACGC TACCTAAGTT TTTGCGGTTG AAAAGTACAT ACCATACTAA	5280
	ACCTAATGCC ATACCAACTG CTGCAATCAA GAATCCGCCG TGGAAGTTTT TAACATTAAAC	5340
	AAAGTGTTGC AAAATAATAG GTGATAATAA TGCACCCATA TTAAGTACA TATAGAAAAT	5400
45	AACAAAACCT GCATCCATAC GTCTATCATT TTCAGGATAT AAACGGCCAA CGATATTTGA	5460
	AATGTTTGGC TTCATTAAAC CTGAACCAAT AATGATGAAG AACATTGATG TGAATAAGCC	5520
	GATTAATGCA AATGGTAAGC TTAAACAAAT ATGTCCGATA ATAATAAAGA CTGCACCTAA	5580
50	TAAAGTAGCG CCTCTAGTGC CTGTAATTCT GTCAGCAATC CATCCGCCTG GTATTGATGT	5640
	CATATAGATT AATGAACCAT AACTGACAT AATTGACATA GCTGTTGTTT TATCAATTCC	5700
55		

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14078 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

	TGGACTATTA ACGGCGaAGA AGATTTAACG AAATACTTAC AAACCAATGT TGATGGTATT	60
	ATCACAGATG ACCCAGCATT AGCTGATCAG ATTAAAGAAG AAAAGAAAGA CGAAACATAC	120
15	TTCGATCGTT CTATAAGAAT TTTGTTTGAA TAATATAAAC AAAGACCTCT AAAGTTATCA	180
	AGATGATACC TTCAGAGGTC TTTTAAATGT TGCCATCTAT GGGATAGGCA ATCGTTTCAT	240
20	TCGTTTATAT TCATATGACA AGTATTTGTA TGGCAATTTG GCGTCACAAA CACTTACATG	300
	ATTTATTGGT GAATTATTAA TTGTTTGTG AATGCAAAGG GTTAGAAATT GAATTGTAA	360
	TACTTTCTAA TCTTTGTTTC GCTTTAGTCA TTTGATCCAA ATTTTtagTG CGTATAGCGG	420
25	ATTTTGCAAT ATAGTGCACA cTAAATATC GCGTTTTGA AACGCATCTA AATTTAGGTA	480
	CGATAATTTA TTAAAGTCAG TGTTCGTAT TAATTCATGT AATTGATCTA CAAGCGCTTG	540
	ATGTTGATAC GTATGTGATG TAGTTTCAGA TTTGCTTGCT AATTTAATAC CAGTCGTATC	600
30	AAGGAGCGCC GCTTTAATAC CAGCAACTAA ATATGTTTTG ATTTTCATTT GTGTTGTCAT	660
	GCTTTGTTAC TCCTTTGATG TACATTAATC AAAAAAATTA TACACTATTG TATATTGCAA	720
	AGCTAATTAA CTATAACAAA AAGATAGTTA ATGCTTTGTT TATTCTAGTT AATATATAGT	780
35	TAATGTCITT TAATATTTTG TTTCTTAAAT GTAGATTGGG CAATTACATT TTGGAGGAAT	840
	TAAAAAATTA TGAAAAAGCA AATAATTTTCG CTAGGCGCAT TAGCAGTTGC ATCTAGCTTA	900
	TTTACATGGG ATAACAAAGC AGATGCGATA GTAACAAAGG ATTATAGTGG GAAATCACAA	960
40	GTTAATGCTG GGAGTAAAAA TGGGACATTA ATAGATAGCA GATATTTAAA TTCAGCTCTA	1020
	TATTATTTGG AAGACTATAT AATTTATGCT ATAGGATTAA CTAATAAATA TGAATATGGA	1080
45	GATAATATTT ATAAAGAAGC TAAAGATAGG TTGTTGGAAA AGGTATTAAG GGAAGATCAA	1140
	TATCTTTTGG AGAGAAAGAA ATCTCAATAT GAAGATTATA AACAATGGTA TGCAAATTAT	1200
	AAAAAAGAAA ATCCTCGTAC AGATTTAAAA ATGGCTAATT TTCATAAATA TAATTTAGAA	1260
50	GAACTTTCGA TGAAAGAATA CAATGAAC TAAGGATGCAT TAAAGAGAGC ACTGGATGAT	1320
	TTTCACAGAG AAGTTAAAGA TATTAAGGAT AAGAATTCAG ACTTGAAAAC TTTTAATGCA	1380

EP 0 786 519 A2

	GTTGTATCAT ATTATGGTGA TAAGGATTAT GGGGAGCACG CGAAAGAGTT ACGAGCAAAA	1500
	CTGGACTTAA TCCTTGGAGA TACAGACAAT CCACATAAAA TTACAAATGA ACGTATTAAA	1560
5	AAAGAAATGA TTGATGACTT AAATTC AATT ATTGATGATT TCTTTATGGA AACTAAACAA	1620
	AATAGACCGA AATCTATAAC GAAATATAAT CCTACAACAC ATAACATAA AACAAATAGT	1680
	GATAATAAAC CTAATTTTGA TAAATTAGTT GAAGAAACGA AAAAAGCAGT TAAAGAAGCA	1740
10	GATGATTCTT GGAAAAAGAA AACTGTCAAA AAATACGGAG AAAGTAAAC AAAATCGCCA	1800
	GTAGTAAAG AAGAGAAGAA AGTTGAAGAA CCTCAAGCAC CTAAAGTTGA TAACCAACAA	1860
	GAGGTTAAAA CTACGGCTGG TAAAGCTGAA GAAACAACAC AACCAGTTGC ACAACCATT	1920
15	GTTAAAAATC CACAGGGCAC AATTACAGGT GAAATTGTAA AAGGTCCGGA ATATCCAACG	1980
	ATGGAAAATA AAACGGTACA AGGTGAAATC GTTCAAGGTC CCGATTTTCT AACAAATGGAA	2040
20	CAAAGCGGCC CATCATTAAG CAATAATTAT ACAAACCCAC CGTTAACGAA CCCTATTTTA	2100
	GAAGGTCTTG AAGGTAGCTC ATCTAAACTT GAAATAAAAC CACAAGGTAC TGAATCAACG	2160
	TTAAAAGGTA CTCAAGGAGA ATCAAGTGAT ATTGAAGTTA AACCTCAAGC AACTGAAACA	2220
25	ACAGAAGCTT CTCAATATGG TCCGAGACCG CAATTTAACA AAACACCTAA ATATGTTAAA	2280
	TATAGAGATG CTGGTACAGG TATCCGTGAA TACAACGATG GAACATTTGG ATATGAAGCG	2340
	AGACCAAGAT TCAATAAGCC ATCAGAAACA AATGCATATA ACGTAACAAC ACATGCAAAT	2400
30	GGTCAAGTAT CATACGGAGC TCGTCCGACA TACAAGAAGC CAAGCGAAAC GAATGCATAC	2460
	AATGTAACAA CACATGCAAA CGGCCAAGTA TCATACGGAG CTCGTCCGAC ACAAACAAG	2520
	CCAAGCAAAA CAAACGCATA TAACGTAACA ACACATGGAA ACGGCCAAGT ATCATATGGC	2580
35	GCTCGCCCAA CACAAAACAA GCCAAGCAAA ACAAATGCAT ACAACGTAAC AACACATGCA	2640
	AACGCTCAAG TGTCATACGG AGCTCGCCCG ACATACAAGA AGCCAAGTAA AACAAATGCA	2700
	TACAATGTAA CAACACATGC AGATGGTACT GCGACATATG GGCCTAGAGT AACAAAATAA	2760
40	GTTTGTAAC TATCCAAAG ACATACAGTC AATACAAAC ATTACGTATC TTTACAACAG	2820
	TAATCATGCA TTCTATGATG CTTCTAACTG AATTAAAGCA TCGAACAATC GGAAGCATAT	2880
45	TTCTAAATTA TTTATTCAAT ATAGTCTTAA ACATAACATG ACCTAATATA TTAATAACCT	2940
	ATTAAAATAA ACCACGCACA TCTAAGTGAT ATACGACAAT CACAGCAATA ATAATTGCTT	3000
	TAGAAAGTCG TGCCGAACTG GAACTTACAA GTCTAGTTCTG AACACACACT GATGTGAGTG	3060
50	GTTTTCTTTA TTTTAAACAT GAACAATCAG ATAAGTTACT AGCATTAGCA AATATTATTA	3120
	AATCAAAGGG CTTGATTCA TAAAATTTAA AACAAATGATT AAAATTAGAC GTGTAAATGT	3180

55

	TATTTACAC	AGCTTCATTA	ATAAAACGAA	ATTGCTTCAA	CCCGCTTCAA	CTTCAACTGG	3300
	CTTCAACTTC	AGCCTACTTC	ATTCAATAAC	AAAACGAATC	CGCTTCATCC	AAAATCAACC	3360
5	ATTCTAACGC	ACATATTCAA	ATATAGCAGC	TGCACCCATG	CCGACACCAA	TACACATCGT	3420
	AACCATGCCG	TAACGGCTAT	CGGGACGTCT	ACCCATTTC	TTAAGTAAAC	GCGCGGTAA	3480
	CATTGCGCCT	GTAGCACCTA	ATGGATGACC	TAAAGCAATA	GCGCCACCAT	TCACATTCGT	3540
10	ACGTGATATA	TCTAGACCTA	CTTCTTTAAT	AGATGCAATC	GTTTGAGAAG	CAAATGCTTC	3600
	GTTCAATTCTG	ATCAAATCAA	TGTCTTCAAC	AGATAGATTG	CTGAGTGACA	ATACTTCAGG	3660
	AATCGCATAT	GCAGGCCCAA	TACCCATAAT	TTTCGGGTCA	ACGCCTACTG	CCTTAAAAAC	3720
15	AACGAATCGT	GCAATAGGTG	TCACGCCGAG	TTCTTTCACT	TTATCTCCAG	ACATTAAAAC	3780
	TACAAATCCT	GCACCATCAG	AAAGTGGGGC	AGATGTTCCCT	GCAGTCATAG	TGCCGTCAGC	3840
	TTTAAATACT	GTACGTAATT	TGGCTAATGC	CTCCATCGTG	GTGTCAGGGC	GTATAAATTC	3900
20	ATCTTGGTCA	AAGATATTTG	TGTGTACTTT	TGGTCTGCG	TTTGTATATT	CAACTGAGTT	3960
	TACTTGTATT	GGAATAATTT	CATCTTTGAA	CCGACCATCA	CGTTGTGCGT	CATAGGCACG	4020
25	TTGATGACTT	CTGACAGCAT	AAGCATCTTG	ATCTTCGCGT	GATACGTCAA	ATTGGGATGC	4080
	TACATTTTCA	GCAGTTAAAC	CCATAGGATA	TGACGCACCT	ATATCATCAT	ATTGTAAGGT	4140
	TGGATTGTTT	GTGGGCTCGT	TGCCACCCAT	TGGTACGGCA	CTCATCAATT	CAACGCCACC	4200
30	AGCTACAAGT	ATATCTCCTT	GACCAGCCAT	AATTTGATTG	GCTGCAATCG	CGATGGTTTG	4260
	TAATCCTGAT	GAGCAGTAGC	GATTCACTGT	TTGACCCGGT	ACCGTGTGAG	ATAATCCCGC	4320
	ACGCAATGCA	ATCGTTTCGT	CAATGTTTTG	GCCTTGTAAT	CCTTCTGGAA	AAGCCGTACC	4380
35	AACAATGACA	TCTTCAATCA	TATTCTTATT	GAATTTTCCG	TCAATACGTT	TCAATACGCC	4440
	TTGTAATACT	TTGGCTGCGA	CATCATCAGG	TCTTTCGTGG	AATAATGCGC	CTTGCTTTGC	4500
40	TTTCGCTGCG	GCTGAACGCC	CATAAGCTAC	AATGTATGCT	TCTTGATGG	TTATCATCCT	4560
	CTCTTAATGA	CTATCTTTTA	ATTACGTAAT	GGCTTACCAG	TTTTTAACAT	ATGTGCAATT	4620
	CTTTCATATG	ATTTTTTAGA	TTTTAGTAAG	TCAATAAAGC	CAATTTTCTC	CAACGATTGA	4680
45	ATGTAACGTT	GATTGATAAA	TGTATTTCTT	GGTAAATCAC	CACCCGCTAA	AATTGTGGCG	4740
	ATATTTAAGG	CAATATGATA	ATCATGGTCG	CTAATAAAAT	GACCCCGTCT	TTGCGCATCT	4800
	AATTGTCCTT	GGATCAATGC	TTTGAAGTCT	TCACCTAAAG	CGATATATTG	ATGTCTAGGA	4860
50	TTCGGAATAT	AGTTTGTTC	TGCTTCATAT	TTCGCACGTT	TGAGCGCAAC	TTGACACGT	4920
	TGTGCTGTAT	TGAAAATAAT	CGTATCTGTA	TCACGTAAAT	AACCATAACG	ACGTGCCTCA	4980

55

EP 0 786 519 A2

	TGTTTGTGTCAT CAAACTTATG CGATGTGCGT AATATGCGAT CAGCCATTTC TGCAAGGCCA	5100
	CCGCCACTCG GTAATAAGCC AACACCTGCT TCAACAAGAC CGATATATGT TTCACTTGCA	5160
5	GCGACAACAA TAGGTGAGTA AAGTACAAGC TCACAGCCAC CGCCTAAGGC ACGACCTTGA	5220
	ACAGCTGTGA CTACTGGTTT CAAACTATAC TTCAAACGAT TAAAGCTATA ATGTAATTGA	5280
	TCAATTGATT GTGCAACGAC ATCATCTACA AGACCGTCTT CATGCGCCTT TTTCATTAAG	5340
10	AAAAGGTTAG CACCCACACT GAAATTGTGA CCATCTGCAT AAATAACCAT ACTTGTGTAA	5400
	TGGTCATTTT CAGTAAATC AATCGCATCA ACTAACGCAT CGTTGAATTC ATCGGTAATG	5460
	ACATTATTTT TACTTTGTAA TTTCAGTAAC AGTTGATCAT CATGAGTTAC GGAAAGTTTG	5520
15	GCATCACCTT TATCCCAAAG TTCATCTTTT ACGAAGTGAG AAATAGGTGT TGCATATTCA	5580
	ATGGTCTCAT CTGTTTATA AAAGCCACCA TCTAAATCAC TAATCCATTG TGGTAAGTCT	5640
20	CCAAGTTCGT CTTCCATACG TGTTTTAAACA CGTTCGTATC CCATTGCATC CCATAATTGG	5700
	AATGGACCAA GTTTCAGTT GAACCCCCAG ACAAGCGCAC GGTCTATGTC TCGGAAATCA	5760
	TCGGTAGCTT TAGGTACATT GATAGCAGAG TAATAGAAAT TATTACGTAA TGTCTCCCAT	5820
25	AAAAATAGTC CCGCTTCGTC TTGCGCATG AATATGGTAT CAAGGTTATG CACTAAGTCT	5880
	TTATTAAATT CATTTAAAT TGGTAATTGT GGTTGCGATA CAGGTACATA ATCTTGTTTT	5940
	TCAACATCGT AAACAAGTCG AGCTTTAGTT TCTTTATCCT TTTTGTAATA TCCTTGTTTC	6000
30	GTTTTACGTC CGAGTGCGCC ATTGTCAAAC AACGTATTTA CAATTTTGAC ATCATGAAAA	6060
	TAAGGTGTTT CTTCAGGTAC TTGTTGCATG CCTTTAATTA CAGACACTGC AATATCTAAA	6120
	CCGACTAGGT CAGATAGCGC ATATGTACCT GTTTTAGGAC GACCAATCGC TTGCCCAGTT	6180
35	AAAGCATCCA CATCTACAAT GCTTATCTTG TGTTGCTCGG CGCGATACAT AATATCATTC	6240
	ATTGTTTGCG TGCCGACTCT ATTTGCGACA AAGCCAGGCA CATCATTGAC GACAATGACA	6300
	CCTTTACCTA ACACATTTTG CGCGAAATTT TTTACATCTA ATATAATAGA TTCCTTCGTG	6360
40	TGTGACGTAG GTATTAACTC CACTAATTTT ATAATACGTG GTGGGTAAAA GAAATGTAGA	6420
	CCAAAGAATC GTTCTTGATC CTTCTCGTTA AATGCTTGAG CAATCGCATT AATTGGAATA	6480
45	CCTGATGTAT TTGTAGCGAA TAAAGCATCT TCTTTAGCAT GTTGTAGAAC TTGTTGCCAA	6540
	ACAGCATGCT TAATTTCAAT ATCTTCTTTG ACTGCTTCGA TATATAAATC AGCATCATCA	6600
	TTTACCAAGT CATCATCAAA ATTACCATAT GTTAAATGAC TCGCTAGATT TAAGTCGAAT	6660
50	AGTAGCGGCC GTTCTTATC TGTAATTTTA TCGTAAGATT TTTTCGCAAT GAGATTTGGA	6720
	TCGTTTTTGT CCACTACAAT ATCTAATAGT TTTACTTTAA GTCCAGCATT CACAAAAAGT	6780

55

	GTGATTCTC CAATTTAGTT GAGGATAAGA TAACCATTAA GATAATTGGA ATAACGTTGC	6900
	TATTTTATAA AATTAATTAA GTATCTTTGA CAGTCATCTT AGCCTCTTAT TTAAGGAAAA	6960
5	AGCTTTATGC TTAAAATAAG TCTTTTTTAG TGAAATTAAT GCATCTCATA TAATTATTTG	7020
	CTATTTATAC GAAAGCAGAA TCTCCAGTCA AAGCGCGTCC AATTACTAAG GCATTAATTT	7080
	CATGTGTACC TTCGTACGTG TAAATCGCTT CTGCATCAGA GAAGAAACGT GCAATATCAT	7140
10	AATCGTCAGC TAGTATGCCA TTACCACCTG TAATACCGCG GCCCATAGCT ACTGTCTCAC	7200
	GCAAACGTAA GGCATTTCATC ATCTTCGCCG TTGAAGTTGC AACCTCGTCA TATTCACCAT	7260
	GTGCTTGCAAT ATTAGCTAAT TGAGCACATG TTGCCATTGC TTGAGCTAAA TTACCTTGCA	7320
15	TCATTGCTAG CTTTTCTTGT ATTAAGTGAT ATTTACTAAT TGGTTTGCCG AATTGCTTAC	7380
	GCTCAGTGAC ATAATCTAAT GTGGCACGTA AAGCGCCAGC CATACCACCT GTAGCCATAT	7440
20	AAGCAACGCC TGCTCTCGTT GAATAAGAA TTTTGCCAAT ATCTTTAAAG CTGTTTATGT	7500
	TTTGTAAGCG ATCCGCTTCA TCTACTTTGA CATTAGTTAA TTAAATTAGG GCGTTAGGAA	7560
	CAATGCGAAG TGCGATTTTA TTATCAATGA CTTCAATATC GACGCCATCT TGTTCGGTC	7620
25	TGACTACAAA GCAATGGGGT TTGCCAGTTT CTTTATTTAC TGCGAATACT GGAATGACAT	7680
	CAGATACATG TGCACCACCA ATCCATTTCT TTTACCATT GATAACCCAA GTATCGCCTT	7740
	GGCGTTCAGC GACTGTTTCA AGACCTCCCG CAACGTCCGA ACCGTGTTCT GGTTCAGTTA	7800
30	AAGCAAAGCA TGTACGCAGT TCATGTGACT GTAATTTAGG TACATATTTT GCAATTTGTT	7860
	CTTTGCTACC TCCGAAATAG AAAGTGTAT GCCCTAAACC TTGGTGAACA CCGAGTAGGG	7920
	TAGCTAAGGA AATATCAAAT CGCGCGAGTA GGTAAGACAT GAAAACTGA AATAGTTGAC	7980
35	TAGGCATTTT GCGGTTTGGA CGATCCTTGT AAAGTAATGG ATTGTTAAAA TAATTTAATT	8040
	CTCCAGATC TTTAAAATAG TCCTCGGGTA CAGTAGCGTC TATCCAATGT TGATTAATAT	8100
	TTTCACGGTA CTTACTTTCT AGCAATGAAT CTACTTGTTG TAAAAATTCG ACTTCACCGT	8160
40	CTGTTAAACC TTTAGCAATA CTAAGTACAT CTTCAGGAAA TAATGTTTTT AAGACCGTTT	8220
	CTTTTCAAA TGTCATATAA ATTCCTCCTA AAAATAATAT GAATACTAAT GTGAAATGCA	8280
45	TTTAATTCAA AAACAACACG CTTTATTTGT AAACGCTTAC ACTAAATGTC AAAAATTTTT	8340
	ATCACCTTTA AAGTGTTTGC GAGACTTTGT CATTATCAT TTGTCGAATC GCAAGTTTAT	8400
	CTGGTTTCTG CGTACTGTTT AACGGCATAT GTGTCACTGG TACATACATT CTTGGGACTT	8460
50	TATAACCTGC TAAACGACTT CGCATATGTT GATTTAAAT TTCAGCGTAA TGAGGTTTAT	8520
	CTTCGCGAAG TATAATGGCT GCAGCAATTG ATTCACCATA TTTTGATGA TCATAGCCAA	8580

55

	AGACATTTTC	GCCACCAGTT	ATGATTAATT	CTTTTTTGCG	GTCAATAATA	AATATATCGC	8700
	CATCGTTGTC	CATCTTCGCT	AAGTCACCAG	TTAATAAATA	TCGACCATGA	AATGCTTTGG	8760
5	CAGTCTCTGC	TGGTTTATTC	CAATATCCTG	GCGTGACATT	TTTAGCCTTA	ATTGCAAGTT	8820
	CGCCAATCTC	ACCAGTAGGT	ACTTCCTCAC	CGTTATCATC	AAGGATACGT	GCATCAACGA	8880
	ACATGACTGC	TTTACCAATA	CTCATTGGCT	TACGTTTTGA	ATTTTCCGGT	GTATTAAACAA	8940
10	GTACAAGAGG	TGCTTCAGTT	AAACCATAGC	CGTTAATAAT	GTTTATGCCA	TATTGTTTAA	9000
	AAGCTGCTTG	GATACTTGGT	AATGGTTGTG	AACCACCTTG	GATGATATAA	TCCATAGCTC	9060
	TAAAAATTTTC	AGGATTAAAA	TTACTAGCAC	GTAGCGTACT	ATAATACATT	GTCGGAATCA	9120
15	TGATAATAAA	TGTAGGGTGA	TATTGTGCAA	TCATGTCATT	CAATTCCTCG	CCGTAAAGT	9180
	AACGTGAAG	AATAAGTGTG	CCACCTGACA	TTAATACTGG	TAATACAGTA	TCGTAAACC	9240
20	CTAAAACATG	GAACATTGGT	GTTGATACAA	TCGTAATATA	GTTTGAATTG	AACTTATACG	9300
	TCAGCTCTAA	GTTTGACCG	TTATGAACAA	ATGATTGATA	TGAGAACATC	ACACCTTTAG	9360
	GTGATCCGGT	TGTACCACTT	GTATAAATTA	ATGCTGCAAG	ATCTTGTTGGT	TCAACAGGTG	9420
25	TTGCTTGAAA	AGGTTGGTGA	TAATCTGGAT	TTACGATTTT	ATCATATTGC	GCCACATCAA	9480
	TATCCATATG	CAATAAGTTT	TGGTCAATAT	CGGTGAGTGA	ACTTAAATGT	TTTTCAGCAT	9540
	AGAAGAGCAG	TTTTAATTGT	GCATCTTCCA	CAATGGCTGC	AATTTCTTTT	GGGTAAAGCC	9600
30	GCCAATTCAA	TGTAATAAAA	ACCGCACCTG	TTTTAAACA	AGCAAACAAT	AAATCTAATA	9660
	TTGCAATATC	ATTTGGCGCA	AAAATACCGA	TAACATCGCC	TTTTTTAACA	CCTTGAGATG	9720
	TTAAATAATG	TGCCATATTA	TCAGCGCGTG	CATTGAGTTG	TTGGTATGTC	CAAGATGTTT	9780
35	GTTTTGCGTG	ATCAATAACG	GCAGGCTTGT	CATCATCGAA	GTCTGAACGC	GTTTTTATCC	9840
	AATCGAAATT	CATTAGTATA	CCCCCTTTAG	CTTCACCTTC	ATACTTTATG	AATTGATTGT	9900
	TTAAGTTGTC	CCCATTTTTT	TTTGTAATG	CTGGTATCAA	TTAATTTTAA	ATGATCAGCA	9960
40	ATAATTGGTT	TAAAAGCCAT	TTGATTCAAA	ATATCTTTAT	GCAAATCAAG	ACCTGGTGCA	10020
	ATTTCAATTA	GTTTCAAGCC	TTGATTGGTG	AGTTCGAATA	CTGCACGATC	AGTAACAAAA	10080
45	TAGATTTCTT	GCTCGAGTGA	TTGTGAATAT	TGTGCATTAA	AGTCGATATG	GCTCACATCT	10140
	GATACAAATT	TCTGGTTTTG	TCCTTCAGTT	TCAATGTTTA	ATCGTTGATT	ATGGCATGAG	10200
	ACATGACTGC	CAGCTACAAA	AGTACCTGAA	AAGATAATTT	TATTTACAGA	TTGCGTAATG	10260
50	TCTATAAAGC	CACCACATCC	ATTTAGTCGG	TCATTGAAGT	AAGACACGTT	GACATTGCCG	10320
	TATTGATCAA	CCTCAGCAAA	GCTAAGATAG	GCAACTGATA	CACCATTGTT	ATAAATAAAA	10380

55

	CGACTCCCAA CGAATCCACC GAAAATGCCA ACATCTAAAA TCGGTTGCAC ATCATGTTCA	10500
	ACACATTCTT CATGCAATAA ATTAGAGAGT TCATTATTGA TGCCATAACC GATGCTAATT	10560
5	GTATCGCCAT AAGTTAAAAA CTGAGCAGCA CGTCGGAGAA TCAATTTGCG ACTATTAAAA	10620
	GGTAATGCGG GTTCAGGTAT TCCATCAATT CGTTCTTCTC CAGACAAGGC TGGTAAATAA	10680
	TGACTTTGAA TTACTTGGCG GTGATTCTTT TCATCTTCTG TGACGTATAC ATAATCGACA	10740
10	AGATTTCTTG GGATAACAAC TTCATTCCGT TTTAGTTGAT AGTCGTCAAC TAAAGCTTTA	10800
	ACTTGTACAA TAACTTTCCC ATGATTGGCT TTCGCGTTTA ATGCGACATG ATAACACTCG	10860
15	CTCAAGTACG CTTCTTGAGT TAAATAAATG TTACCTTGTT GATCTGCGTA TGTTCCTCTC	10920
	AGTAGTGCCA CATCAACGCT AGGGAATGTG TAATGTAAGT ATGTTTCATC GTTGATGGTT	10980
	ACTAATGAAA CTAAATCATC CGTTGTTTCG GTATTTACTT TACCGCCACC GTATCTAGGA	11040
20	TCAACAGCTG TGTTTAATCC GATTTTAGTA ATAACTCCAG GTAATAATTG ATTACTCTGA	11100
	CGATAATGAG TTGCAATGAT ACCTTGTTGGT AAAAAATAAG CTTCAATGTC ATTATTTTTC	11160
	ATTGcTTGTG CCGTTTTGGA AGAAGCCGTT AAAATACTCA TAATGACACG TTTAATCATG	11220
25	CGACGTTCTA TAAATCATC TAAATCCGGT GCGGCACCTA AACTATGAAT ATCATTCGCT	11280
	AATATAAACG TTAAATCATT GGGCGTATGA TATGTGTCAT GTTGCGCTAA CACAGCACGT	11340
	AGAACTTCGG CGGGTAAGTT GGCTACAGCT AATGCTGGTA AACCAATCAC ATCACCATCT	11400
30	TTAATGATAT GTTGTAAGTC GTGCCATGTG ATTTGTTTCA AGCAAGTCAC CTCCATCACA	11460
	TTTGATAAAA TATAGCGTTT TTACACTTTG TGTAACCCCT TaCAAGAAAT ATAACATAAC	11520
	GACGTTTAAA ATCAATTAGA AATATCTTTT TATTCTGATA ATAGACACAG TATAGACACA	11580
35	TTTTGATGGT CGATAACAAT TGTAATATCA AGGGTTTGTA ATGAATTGAA TATCATTAAG	11640
	ATACTTATAT AAAAATATTG TTCGGAATAT AAAAAGTTAA ATAGGTTTTG ATTTTAAAT	11700
	ATGAAATACA AAGTGCCCAA TCGAACAAAG TATTTATATT AAAATATGGA AAATCCATCA	11760
40	ATATTAAATT AAAATAGTTT TATTATGAAA AGTGAAAGTA GGTAAGTCTA TGGAAGGTCT	11820
	TAATCATCGA AGAAATACAG AAAAAGAAGA GACAACACAA ACGCAATCaG TTGCACCTAA	11880
45	TACAGGTGAA GAGGGGATGT CATCAGCAAG TACACAATCA ACTAAGACGT CCGACATACA	11940
	TAATGAATCT ATCGATAAAC AAATGGAAGC TAAAGCGCAT GAAACAGCGC AAAATACAGA	12000
	TTTAAAAAAC GAAGCAAGAA GTTTATTTGA TAATGCAACC AAATCAATCG GTAGACTAGC	12060
50	GGGCAATGAT GAAAGCTTAA ATCTTAATTT AAAAGATATG CTTTCTGAAG TATTTAAGCC	12120
	GCATACTAAA AACGAAGCAG ATGAAATATT TATAGCGGGT ACTGCTAAAA CTACGCCAGC	12180

55

	TTTCACAGTA ACATTATTG GATTATGGGT CATGGCAGCA ATTTTAAATA AACTAACGC	12300
	GATTCCGGGT CTCATTTTTA TAGGGCTTT AACAGTACCA TTATCGGGTT TGTTCTTCTT	12360
5	TTATGAATCA AATGCGTTTA AAAATATTAG CATTTTGTAA GTTATTATCA TGTTCTTTAT	12420
	TGGCGGCGTA TTTTCATTAC TAAGTACGAT GGTATTATAT AGATTGTGCG TTTTATAGTA	12480
	TCAATTCGAA AGGTTTGGTT CTTTAACATT TTCGATGCA TTTTATAGTAG GATTAGTTGA	12540
10	AGAAACTGGA AAAGCACTCA TTATGTGTTA TTTCGTCAAT AAATTGAAAA CAAATAAGAT	12600
	TTTGAATGGA TTATTAATCG GTGCTGCTAT TGGTGCAGGG TTCGCAGTTT TTGAATCAGC	12660
	AGGTTATATT TTGAATTTG CTTTAGGAGA AAATGTCCCA TTATTAGATA TTGTCTTCAC	12720
15	ACGTGCGTGG ACTGCGATTG GTGGTCATTT AGTTTGGTCA KCGATTGTTG GTGCTGCAAT	12780
	AGTTATTGCG AAAGAACAGC ATGGCTTTGA ATTCAAAGAT ATTTTGTATA AACGCTTTTT	12840
20	AATATTCCTT TTATCAGCCG TTGTTTTACA TGGCATTGG GATACATCTT TAACTGTACT	12900
	TGGCAGTGAT ACGTTGAAAA TATTTATTTT AATCGTTATT GTGTGGATAC TTGTATTCaT	12960
	TTTAATGGGG GCAGGTTTAA AACAAAGTAA TTTACTGCAG AAAGAATTTA AAGAACAACA	13020
25	GAAAAAAGTA GACGAATAAT AATTAAAGCT TATGTTGCTC ATATGTTTGT GACATAAGCT	13080
	ATTTTATAA TTTGTCTTTA AAAGAGTGGG ATAGGAATAC TTTTGGAGT TAAAAAGTG	13140
	TTtCACGTTA AACAAATAGT GACAATTAGA TTTATATAAA ATGAACATGA TTCACTGAAA	13200
30	GTATGTAATA ATCATTTTAT TGAAATTCAT CAAACAGAAA TTAATACAAT CATATAAGCA	13260
	AATTAAACCA CGCCATAATC ATATTGGATG ACTTCGGCGT GGTTTTTATA GTTGAAGCAG	13320
	GGCTGAGACA TAAATCAATG TCCCACACTC CCTTATCGTT CAATCGTTGT TCGATAATCG	13380
35	ATTAAATAGA TACCTTCAGG TGTIACCTTA TAATTTTAA CCTTAGAGTT AGCAGCGACT	13440
	ATTGATCGT TGTAAGCAAT ATAAGTGTG GGTACATCTC GACTTGATAA TTTAATAATA	13500
	TCATTAGAAA TATTGTGACG TTCCTTAACA TCTACAGTAT GATTCAATTG ATTAATTAAA	13560
40	TCATCGACGT TGCTATTATT GTAGTCTCCT TTATTAATAG CACCATCTTT TTTATATGCT	13620
	TGATTAAAGA AATAACCTGT ATCTCCACGA GGAATTGTTT CGAAACTATA CATCGTTGCA	13680
45	TCCCATGCAG AACGGTCTTT TAAGTAACCT TCTATGTCAT CAACACTTTT AATGTCGATT	13740
	TCAATATTG CTTTTTATAG ATCTGATTGT AATACTTGCG CAATTTTCGA TAGCTCTGGA	13800
	CGACCGTCAT ACGTAATTAA CTTAATTTTT AAAGGGTGTT CTTTTGTATA ACCATCTTTA	13860
50	GCTAATAACA TTTTGTCTG TTCGATATTT TGTTTGGTTA ACTTAGGTTT TTTAATATAT	13920
	GGAATTTTAT CATTAAATGG ACTCGTTGCA GGTTTCGCAT AACCTTGATA AATATGATCT	13980

55

TTATTAGTAT GATTATACAT AAGTaAGAAG TTCTAAAn

14078

(2) INFORMATION FOR SEQ ID NO: 192:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 486 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

TGAAA	ACTAA	AGTGT	TTCTA	ATGCG	TGACT	AAAAT	TAGTA	ATAAT	TAGT	TCTCA	TGATA	60
ATAGG	TATTT	TTGAA	AAATG	GAGGA	GTCTA	TAAAT	TGGTA	AAAAA	TGGG	TCTAG	GTTTA	120
TCTAT	TGCAT	TGGTT	GTTAT	TGGTA	TGCCC	GTTGT	TATGT	TAATG	ATTTT	TTCTA	GTCAA	180
AAAAC	GACTT	ATTTT	TGGTT	A	TATGA	ATAGT	AATAC	AAATG	CAGAA	AAAGT	TGTCAG	240
AAAGAT	GGAT	TAGTCA	AAACA	TAATAT	CAAA	GTAGA	ACCAT	CTAAT	GATTT	CAAGC	CGAAA	300
AAAGG	AGACT	TTGTA	AAATT	AGTTT	CTAAA	GATGAT	TGGGA	AGACAT	TTTTA	TAAAC	AAGAG	360
ATTGT	TAAAC	ATGAT	GACGT	CCCAC	ACGGT	TTAAT	GATGA	AAATTC	CACGA	CATGC	CATATG	420
AATTA	ATAAAA	AAAGCAT	CTA	TAACG	TAAAT	TTGAAG	AAGT	AGAGT	TATCT	TCTTA	TGCGT	480
TTTAGA												486

(2) INFORMATION FOR SEQ ID NO: 193:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1626 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

GAGGT	CTATA	TACAAT	TATG	GTTGT	TCCAG	TTAAAC	GAAC	TGATG	GCCTT	ATTACT	TAAGT	60
TTAAT	AGATT	AATTGA	AAGA	CGATT	TATTAC	GTCATT	TCAG	TAAAA	AAGGT	TATAT	CACAT	120
GGGAG	GAAAA	TTGATT	TGCT	GACATT	TTTAA	AATGT	ATCGG	TTGTG	GTGCG	CCACT	TCAAT	180
CTGA	AAGATA	AAATA	AACCT	GGTTT	TGTAC	CAGAG	CATAA	TATGT	TTTCGT	GATGAC	GTGA	240
TTTGC	CAGACG	TTGTT	TCCGC	TTGAAA	AAATT	ATAAC	GAATT	CAAGAT	GTAG	GATTAG	AAAG	300
TGA	AGACTTT	TTAAA	ATTAT	TATCAG	GAAGT	TGCGG	ATAAA	AAGGG	TATTG	TCGTCA	ATGT	360
CGT	GGATGTA	TTGACT	TTTG	AAGGAT	CATT	TATTA	ATGCA	GTAAAC	GTGTA	TTGTC	GGAAA	420